HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2020 September 08.

Published in final edited form as:

Neuroimage. 2020 September ; 218: 116878. doi:10.1016/j.neuroimage.2020.116878.

Anterior superior temporal sulcus is specialized for non-rigid facial motion in both monkeys and humans

Hui Zhang[a,b,*,1], **Shruti Japee**[a,1], **Andrea Stacy**[a], **Molly Flessert**[a], **Leslie G. Ungerleider**[a]

[a]Laboratory of Brain and Cognition, NIMH, NIH, Bethesda, MD, 20892, USA

[b]Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing, 100191, China

Abstract

Facial motion plays a fundamental role in the recognition of facial expressions in primates, but the neural substrates underlying this special type of biological motion are not well understood. Here, we used fMRI to investigate the extent to which the specialization for facial motion is represented in the visual system and compared the neural mechanisms for the processing of non-rigid facial motion in macaque monkeys and humans. We defined the areas specialized for facial motion as those significantly more activated when subjects perceived the motion caused by dynamic faces (dynamic faces > static faces) than when they perceived the motion caused by dynamic non-face objects (dynamic objects > static objects). We found that, in monkeys, significant activations evoked by facial motion were in the fundus of anterior superior temporal sulcus (STS), which overlapped the anterior fundus face patch. In humans, facial motion activated three separate foci in the right STS: posterior, middle, and anterior STS, with the anterior STS location showing the most selectivity for facial motion compared with other facial motion areas. In both monkeys and humans, facial motion shows a gradient preference as one progresses anteriorly along the STS. Taken together, our results indicate that monkeys and humans share similar neural substrates within the anterior temporal lobe specialized for the processing of non-rigid facial motion.

Keywords

Facial motion; fMRI; STS; Emotion processing

[*]Corresponding author. Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University, 37 Xueyuan Road, Haidian District, Beijing, 100191, China., hui.zhang@buaa.edu.cn (H. Zhang).

[1]Hui Zhang and Shruti Japee contributed equally and are co-first authors.

CRediT authorship contribution statement

Hui Zhang: Conceptualization, Software, Writing - review & editing. **Shruti Japee:** Conceptualization, Software, Writing - review & editing. **Andrea Stacy:** Investigation, Data curation. **Molly Flessert:** Investigation, Data curation. **Leslie G. Ungerleider:** Supervision, Writing - review & editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2020.116878.

1. Introduction

Facial motion, a special type of biological motion, transmits a wealth of information for effective social interaction and communication. For example, both humans and other primates depend on facial expressions to convey emotion to others during social interactions. Although abundant evidence has shown that motion enhances the recognition of facial identity (Knight and Johnston, 1997; Lander et al., 1999; Knappmeyer et al., 2003; Roark et al., 2006) and facial expressions (Wehrle et al., 2000; Ambadar et al., 2005; Trautmann et al., 2009, see O'Toole et al., 2002 for review), the underlying neural mechanisms mediating the perception of facial motion are still unclear.

The results of several previous studies have suggested the existence of specialized neural substrates for processing facial motion in primate superior temporal sulcus (STS). In macaques, an fMRI study found that a region deep within the fundus of anterior STS showed significant greater activation evoked by facial motion but not by dot motion (Furl et al., 2012). By contrast, Polosecki and colleagues (Polosecki et al., 2013) examined the interaction between faces and motion within predefined face-selective patches and found that facial motion is not represented in the fundus of STS but rather within the anterior lateral (AL) and anterior medial (AM) face patches. However, Fisher and Freiwald (2015a), who also examined face and motion interactions within the predefined face-selective patches reported distinctive facial motion selectivity in dorsal bank STS face patches. In humans, Fox and colleagues (2009) showed that a dynamic face localizer (contrasting activations for dynamic faces with dynamic objects) engaged more areas within the STS than a static face localizer. Pitcher and colleagues (Pitcher et al., 2011) reported that the STS responded more strongly to dynamic than to static faces, and this preference for dynamic faces was most striking in the right anterior STS.

Taken together, these studies have suggested a role for face-selective regions within STS in the processing of dynamic faces in both monkeys and humans, but they have left a key question unanswered. Are there common mechanisms in monkeys and humans that mediate the specialization of STS for facial motion processing? Only one study so far (Polosecki et al., 2013) has attempted to characterize facial motion selectivity along STS in both species in an equivalent manner. This study reported that in monkeys only face patches AL and AM encode facial motion better than object motion, while in humans, only pSTS (referred to as STS-FA by Polosecki et al., 2013) showed a preference for facial motion. This study concluded that facial dynamics were analyzed via different mechanisms in the two species. This conclusion is at odds with work done in the two species separately that seems to suggest a posterior-anterior gradient in facial motion preference in monkeys (Fisher and Freiwald, 2015a) and in humans (Pitcher et al., 2011).

The conflicting results regarding monkey STS face patches between Polosecki et al. (2013) and Fisher and Freiwald (2015a) may be due to the inconsistent definition of facial motion selectivity between the two studies. This inconsistency also exists in the definition of facial motion preference across the two species (Pitcher et al., 2011; Fisher and Freiwald, 2015a; Polosecki et al., 2013). Further, most human and monkey studies investigating facial motion selectivity have been limited to predefined face-selective ROIs within STS (Pitcher et al.,

2011; Polosecki et al., 2013; Fisher and Freiwald, 2015a) which provides limited information about the distribution of facial motion preference outside of face-selective ROIs across both species. Additionally, to our knowledge, none of the previous studies precisely controlled the amount of motion in the stimuli - a factor that can affect the strength of fMRI activation in facial motion preferring regions.

Considering the above issues, it is important to systematically examine facial motion selectivity across the entire STS in both species, using improved methodology. To this end, we explicitly defined facial motion selective regions as those significantly more activated by motion caused by faces (dynamic > static faces) than motion caused by non-face common objects (dynamic > static objects). Critically, and unique to our study, we ensured that the motion energy was equivalent in dynamic faces and dynamic objects using an optic flow algorithm (Beauchemin and Barron, 1995; Fleet and Weiss, 2006; Lucas and Kanade, 1981; Baker and Matthews, 2004; Bruhn et al., 2005). We then obtained detailed facial motion selectivity maps across the entire visual system of four macaque monkeys and sixteen human subjects. Further, we elucidated the motion preference for faces and objects in each examined region of interest within and outside STS. Overall, the aim of our study was to systematically uncover the representation of facial motion within the visual association areas in both monkeys and humans, with the goal of gaining insight into common neural mechanisms across the two species.

2. Materials and methods

2.1. Subjects

Monkeys: Four male macaque monkeys (Monkeys F, R, K, and B; *Macaca mulatta*, 7.6–10.9 kg; 5–8 years) were used. All animal procedures followed the Institute of Laboratory Animal Research (part of the National Research Council of the National Academy of Sciences) guidelines and were approved by the National Institute of Mental Health (NIMH) Animal Care and Use Committee. Each monkey was surgically implanted with an MRI-compatible head-post in sterile conditions under isofluorane anesthesia. After recovery, monkeys were trained to sit in a sphinx position in a plastic restraint chair (Applied Prototype) and to fixate a central target for long durations with their heads restrained, facing a screen on which visual stimuli were presented.

Humans: Twenty-three human subjects (11 male) aged 24.0± 4.7 (mean ± SD) years participated in the study. All subjects were right-handed, had normal or corrected-to-normal vision, and were in good health with no past neurological or psychiatric history. All participants gave informed consent according to a protocol approved by the Institutional Review Board of the National Institute of Mental Health. Data from seven subjects were discarded because of excessive head movement (>4 mm) in the MRI scanner during fMRI data collection, leaving a total of 16 subjects (7 male), aged 25.4 ± 4.2 (mean ± SD) years, for further analysis.

2.2. Stimuli

Dynamic and static stimuli showing monkey faces, human faces, and non-face objects were used.

2.2.1. Video materials—Dynamic monkey face videos were recorded from more than 20 unfamiliar male lab macaque monkeys, with their heads fixed, seated in a vertical chair and facing the camera. The videos were taken while the monkeys blinked, moved their eyes, or chewed, but not when they made typical emotional facial expressions such as fear-grin, open-mouthed threat or lip-smack. Dynamic human face videos were recorded from 10 male and 10 female human actors, who kept their heads as still as possible during recording and moved their eyes and face muscles without expressing any emotion. Dynamic object stimuli were all naturalistic inanimate moving objects, including fluttering flags, jumping flames, whiffling leaves, rotating gears, ticking clocks, moving cars, etc. Most video clips of moving objects were downloaded from free online websites, while other video clips were created expressly for the purposes of the current experiment. All the dynamic object videos depicted non-rigid motion. For example, the moving car video showed a model car with its tires rotating and the windshield wipers moving back and forth while the body of the car remained still. This kind of articulated motion (i.e., individual rigid parts of an object moving independently of one another) is considered to be a type of non-rigid motion (Aggarwal et al., 1998). Further, the video clip of a moving plane (that was used only in our human fMRI experiment) contained a mix of rigid motion (body of the whole plane rotating) and non-rigid motion (aircraft navigation lights flashing). Similarly, our human face video clips contained very small amounts of uncontrolled rigid head motion, and thus the non-rigid motion in the object stimuli and human face stimuli were balanced to a substantial extent.

The dynamic videos were initially recorded at a resolution of 1280 × 720 pixels and 29.7 frames per sec (fps). Using Adobe Premiere Pro CC, the videos were cropped to show only the faces or objects on a black background, and were resized so that the moving stimuli in the video clips subtended roughly the same field of view. These videos were further processed in MATLAB, to normalize the luminance, by using the following formula to make all the video clips have a mean value of 80 and variance of 2500:

$$2500 \times (V\text{-}\mathrm{Mean}(V))/\mathrm{Variance}(V) + 80$$

where, V is one video clip containing 60 image frames, Mean(V) is the mean, and Variance(V) is the variance of all non-zero gray-values in video clip V. The videos were then converted to gray-scale, and trimmed to 2-sec clips. These above edits in terms of size, field of view, luminance and contrast were to keep the low-level features as similar as possible across videos.

2.2.2. Motion energy evaluation—Motion energy contained in the video clips was evaluated using an optical flow algorithm (Beauchemin and Barron, 1995; Fleet and Weiss, 2006) and the Lucas–Kanade method (Lucas and Kanade, 1981; Baker and Matthews, 2004; Bruhn et al., 2005), which were implemented with in-house C++ and MATLAB code (See Supplementary Material A. and Figure S1 and S2). Supplementary Figures S1 and S2 show

the calculation of the optical flow velocity fields and the evaluation of motion energies in a dynamic monkey face video clip and a dynamic non-face object video clip. The optic flow was defined as the pattern of apparent motion of the object between two consecutive image frames caused by the movement of the object. The optical flow velocity field was evaluated for each pair of consecutive frames in each stimulus video clip. Motion energy was calculated as the sum of the absolute values of the velocity across all the spatial pixels of the optical flow field and across all the consecutive optical flow velocity fields. It should be noted that this optical flow algorithm can only evaluate the motion energies of video clips taken with fixed cameras, and not the motion energies of video clips taken with cameras moving during video filming.

2.2.3. Experimental stimulus creation—We created more than two hundred 2-sec video clips for each category of dynamic monkey faces, dynamic human faces, and dynamic non-face objects as a candidate stimulus set. We then selected pairs of dynamic monkey faces and dynamic object video clips, and dynamic human faces and dynamic object video clips from the candidate stimulus set. These pairs were chosen such that the motion energy in the face video was equivalent (or as close as possible) to that in the corresponding object clip. We selected 12 such pairs of dynamic monkey faces and their corresponding dynamic object clips, as well as 10 pairs of dynamic human faces and their corresponding dynamic object clips. The corresponding static human face, monkey face and object stimuli were generated by extracting the most representative frame from the dynamic video clips.

2.3. Monkey fMRI experiments

2.3.1. Main experiment—Four monkeys viewed blocks of four stimulus categories: dynamic monkey faces, dynamic objects, static monkey faces and static objects. Each dynamic category contained 12 different video clips, and each static category contained 12 different images that were generated from the corresponding dynamic video clip. Each block lasted 24 s, with a 16-sec baseline fixation period between blocks. At the beginning and end of each run, baseline blocks were presented for 12 s and 20 s, respectively. Within a block, each video clip (for dynamic categories) or each image (for static categories) was presented for 2 s, with no blank periods between them (see Fig. 1A). There were three blocks for each condition per run, with a random ordering of the block within a run. Each run lasted 8 min 16 s. Monkey F completed 36 runs in three sessions, monkey R completed 40 runs in three sessions, monkey K completed 41 runs in three sessions, and monkey B completed 18 runs in two sessions.

2.3.2. Localizer—All four monkeys also performed traditional localizer runs by viewing blocks of static neutral monkey faces, static objects and static scrambled monkey face images, to identify each monkey's static face patches. The objects included inanimate images of eyeglasses, shoes, shirts, scissors, cups, umbrellas, keys, etc. The scrambled face images were constructed by randomly permuting the phase of the face images in the Fourier domain. The images used in the localizer runs differed from those used in the main experiment. Each of the three categories contained a total of 30 different images. These images were converted to gray-scale, normalized to have equivalent size, luminance and contrast, and resized to 350 × 350 pixels. Each block within a run lasted 24 s, with a 16-sec

blank period between blocks. Blank blocks were also presented for 12 s and 20 s, respectively, at the beginning and end of each run. Within a block, each image was presented for 500 msec, with 1-sec blank periods between the images. There were two blocks for each condition per run, with all blocks presented in random order within a run. Each localizer run lasted 4 min 16 s. Each of the four monkeys was scanned in two localizer sessions, for a total of 32 runs collected for monkey F, 36 runs collected for monkey R, 32 runs collected for monkey K, and 27 runs collected for monkey B.

For both the main experiment and localizer, stimuli were presented using Presentation software (version 12.2, www.neurobs.com) at a resolution of 1024 × 768 pixels, and a refresh rate of 60 Hz. Images were displayed via an LCD projector (Avotec Silent Vision SV-6011–2) onto a front-projection screen positioned within the magnet bore, spanning a visual angle of 11 × 8°, centered at fixation. Monkeys were required to maintain fixation on a red rectangle (visual angle: 0.4 × 0.4°) super-imposed on the stimulus center to receive liquid reward. The fixation size was adjusted before every scan session, ranging from 2 × 2° to 2.5 × 2.5 degrees of visual angle, based on each monkey's daily performance and the noise of the eye-tracking system. In the reward schedule, the frequency of reward increased as the duration of fixation increased (Hadj-Bouziane et al., 2008; Bell et al., 2009; Furl et al., 2012; Liu et al., 2013). Eye position was monitored with an infrared pupil tracking system (iView, Inc.). Each animal was scanned for a total of four to five sessions.

2.4. Human fMRI experiments

2.4.1. Main experiment—Human subjects viewed blocks of dynamic human faces, dynamic objects, static human faces and static objects while they performed a one-back working memory task, pressing the left button if the current video clip matched the preceding one, and the right button if it did not. Each dynamic category contained 10 different video clips, with 2 video clips presented twice consecutively in one block. Each static category was generated from its corresponding dynamic category. Each block last 24 s, with a 16-sec blank period between blocks. Blank blocks were also presented 8 s and 16 s, respectively, at the beginning and end of each run. Within a block, each dynamic video clip or each static image was presented for 2 s, without blank periods between them. There were two blocks for each condition per run, in random order within a run, and the order of the blocks was also randomized across runs (See Fig. 1B). Each run lasted 5 min 28 s. Each subject completed 5 runs of the task.

2.4.2. Localizer—Human subjects also performed two traditional localizer runs to identify each individual's static face-selective regions. During these runs, subjects viewed blocks of static neutral human faces, static objects, and static scrambled human faces, and were asked to press the left button if the current image matched the preceding one, and the right button if it did not (one-back working memory task). The images used in the localizer runs differed from those used in the main experiment. Each block lasted 24 s, with 16-sec blank periods between blocks. Within a block, each image was presented for 500 msec, with 1-sec blank periods between images. There were two blocks for each condition per run, with blocks presented in random order within a run. The stimulus set and presentation parameters

were identical to those used in the monkey localizer runs, except that monkey face images were replaced with human face images.

For both the main experiment and localizer, stimuli were displayed with Presentation software and back-projected onto a screen in the dimly lit scanner room using an LCD projector (PLUS U2–1200) with a refresh rate of 60 Hz and resolution of 1024 X 768. Subjects viewed the stimuli via a mirror system installed in the head coil. They were instructed to maintain fixation on a central gray cross at all times during the whole fMRI scan. The face and object stimuli subtended a visual angle of 8.2 × 6.0°, the gray fixation cross subtended a visual angle of 0.2 × 0.2°, in both the main experiment and localizer.

2.5. fMRI acquisition

2.5.1. Monkeys—Before each scanning session, an exogenous contrast agent (monocrystalline iron oxide nanocolloid [MION]) was injected into the femoral or external saphenous vein (12–15 mg/kg) to increase the contrast/noise ratio and to optimize the localization of fMRI signals. Imaging data were collected in a vertical 4.7 T Bruker scanner with an eight-channel surface coil. The functional images were acquired with a single-shot interleaved gradient-recalled echo planar imaging sequence, with coronal slices positioned to cover all the temporal lobe and most of the occipital lobe (TE = 13.8 msec; TR = 2000 msec; flip angle = 90°; matrix size = 64 × 36; field of view: 96 × 54 mm; voxel size = 1.5 mm × 1.5 mm x 1.5 mm; 28 coronal slices, no acceleration was used). Supplementary Figure S3A shows the fMRI scan slice coverage overlaid on a representative monkey brain. A low-resolution anatomical scan was also acquired in the same scan session to serve as an anatomical reference (modified driven equilibrium Fourier transform sequence: TE = 2.932 msec; TR = 6.24 msec; flip angle = 12°; matrix size = 128 × 128; field of view: 96 × 96 mm; voxel size: 1.5 mm × 0.5 mm × 0.5 mm; 48 coronal slices). In a separate scan session, high-resolution anatomical scans were obtained from each monkey under anesthesia in a horizontal 4.7-T Bruker scanner, using the modified driven equilibrium Fourier transform sequence (TE = 4.9 msec; TR = 13.8 msec; flip angle = 14°; voxel size 0.5 mm × 0.5 mm x 0.5 mm). These high-resolution anatomical data were used to create the cortical surface for each monkey using FreeSurfer (http://surfer.nmr.mgh.harvard.edu).

2.5.2. Humans—Imaging data were collected on a Siemens 7T scanner with a Siemens 32-channel surface coil. Each scan session began with a high resolution T1-weighted Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequence (TE = 3.88 msec; TR = 3000 msec; flip angle = 6°; matrix size = 256 X 256; voxel size = 1 mm × 1 mm x 1 mm; 192 axial slices). The functional images were acquired with a single-shot interleaved gradient-recalled echo planar imaging sequence, with slices positioned to cover all of the occipital and temporal lobes (TE = 27 msec; TR = 2000 msec; flip angle = 70°; matrix size = 126 × 126; voxel size = 1.6 mm × 1.6 mm x 1.6 mm; 43 oblique axial slices, no acceleration was used). Supplementary Figure S3B shows the fMRI scan slice coverage overlaid on a representative human brain.

2.6. Data analysis

2.6.1. Preprocessing—For monkeys, only scanning sessions with adequately high behavioral performance (>90% central fixation throughout the duration of each run) were analyzed further. For both monkeys and humans, fMRI data were analyzed using AFNI (Cox, 1996). Data from the first four TRs in human scans and the first six TRs in monkey scans from each run were discarded. The remaining images were slice-time corrected, realigned to the first volume of each session, and spatially smoothed with a 3-mm FWHM Gaussian kernel for both species. Signal intensity was normalized to the mean signal value within each run and multiplied by 100 so that the results after analysis represented percentage signal change from mean. Regressors of interest were created by convolving each stimulus condition with the MION function (for monkey data) or the gamma function (for human data), and then input into a general linear model (GLM) for parameter estimation. The slow drifts and six head movement parameters (roll, pitch, yaw, dS, dL, dP) were included in the GLM as regressors of no interest.

2.6.2. Localization of facial motion selective regions—Data from the main experimental runs (from sessions 1 and 2 in monkeys; Monkey F: 26 runs; Monkey R: 28 runs; Monkey K: 28 runs; Monkey B: 18 runs; and runs 1–3 in humans) were used to define facial motion selective regions for each monkey and human. We defined regions selective for facial motion as those responding more to facial motion (dynamic faces > static faces) compared to object motion (dynamic objects > static objects). We used this contrast to generate the whole brain facial motion selectivity map and thresholded the resulting map at $p < 0.001$ (FDR corrected) to localize the facial motion selective regions.

2.6.3. Localization of face-selective regions—Data from the localizer runs were used to define static face-selective regions for each monkey and human, by contrasting the fMRI response to static faces with that to static objects ($p < 0.001$, FDR corrected).

In addition to the above static face selective regions, data from some of the main experimental runs (same as the ones used for facial motion selectivity) were also used to define dynamic face-selective regions, by contrasting the fMRI response to dynamic faces with that to dynamic objects and thresholding the resulting map at $p < 0.001$ (FDR corrected). We also delineated regions that were sensitive to motion in general (dynamic stimuli > static stimuli), motion in faces (dynamic faces > static faces) and motion in objects (dynamic objects > static objects) using a threshold of $p < 0.001$ (FDR corrected). These maps are further discussed and shown in Supplementary Material B and Supplementary Figure S4 and S5.

2.6.4. Conjunction analysis—To evaluate the overlap between maps of face selectivity and facial motion selectivity, we performed a conjunction analysis at the individual level in both monkeys and humans. For each individual subject, the maps of facial motion selectivity and face selectivity were first thresholded at $p < 0.05$ (uncorrected), and then combined using a logical AND operation to generate a conjunction map (using the AFNI function *3dcalc*). The final statistical threshold for the conjunction map was $p < 0.0025$ ($0.05^2 = 0.0025$, Bonferroni corrected).

2.6.5. Region of interest (ROI) definitions—We defined multiple ROIs that included regions that were consistently activated in the face selectivity or facial motion selectivity contrast maps described above. In both monkeys and humans, the ROIs in visual cortex were selected by drawing a sphere of 6-mm radius around the peak of the activation in face selectivity or facial motion selectivity maps thresholded at $p < 0.001$ (FDR corrected). In monkeys, a ROI in the amygdala was selected by manually drawing a region within the amygdala's anatomical boundaries, around the peak of the face-responsive activation (static face > static scrambled face; $p < 0.001$, FDR corrected). In humans, the ROI for the amygdala was selected by manually drawing a region around the peak of the face-selective activation ($p < 0.001$, FDR corrected) that fell within the amygdala's anatomical boundaries.

2.6.6. Motion sensitivity within ROIs—Since the main goal of our study was to understand how preference for facial motion is distributed within the visual system, we quantified the motion sensitivity for faces and objects separately within each ROI as the difference between the PSC values for dynamic and static stimuli.

2.6.7. ANOVAs within ROIs—Data from separate main experimental runs (from session 3 in monkeys; Monkey F: 10 runs; Monkey R: 12 runs; Monkey K: 13 runs) were used for the ROI analysis in monkeys. The fourth monkey (Monkey B) did not yield any data for ROI analysis because his head-post came loose prior to session 3. fMRI percent signal change (PSC) for dynamic monkey faces, dynamic objects, static monkey faces and static objects from each run were averaged within each ROI and used to estimate motion sensitivity and facial motion selectivity for each region. Similarly, runs 4 and 5 from the main experiment were used for ROI analysis in humans. fMRI PSC data for dynamic human faces, dynamic objects, static human faces and static objects for each subject were averaged within each ROI and then used to estimate motion sensitivity and facial motion selectivity in each region.

In monkeys, the PSC data were submitted to a repeated-measures analysis of variance (ANOVA). We defined category (levels: faces, objects), motion (levels: dynamic, static), hemisphere (levels: left, right) and ROIs as within-subject fixed-effect factors, and Monkey (levels: Monkey F, Monkey R, Monkey K) was defined as a between-subject fixed-effect factor. Defining monkey as a fixed-effect factor is common in monkey fMRI studies (Jastorff et al., 2012; Polosecki et al., 2013; Fisher and Freiwald, 2015a; Sliwa and Freiwald, 2017) where the number of subjects (3 monkeys in this study) is usually too small for effective power in the random-effects analysis. To understand the nature of the interactions, we also performed a category (levels: face, object) x motion (levels: dynamic, static) x hemisphere (levels: left, right) mixed-design repeated-measures ANOVA for each ROI.

In humans, the PSC data were submitted to a repeated-measures ANOVA with category (levels: face, object), motion (levels: dynamic, static) and ROI as within-subject factors. We also performed a repeated-measures ANOVA for each ROI, with category (levels: face, object) and motion (levels: dynamic, static) as within-subject factors. ANOVA was implemented in SPSS (https://www.ibm.com/analytics/spss-statistics-software, IBM SPSS Statistics 24).

3. Results

3.1. Facial motion selectivity

3.1.1. Monkeys—The contrast of fMRI response to motion caused by faces (dynamic faces > static faces) relative to motion caused by objects (dynamic objects > static objects) ($p < 0.001$, FDR corrected) identified areas preferentially selective for facial motion compared to object motion (see Fig. 2A). These facial motion selective regions were found in the fundus and upper bank of the anterior STS for monkeys F, R and K bilaterally, and for monkey B in the right hemisphere. No other brain region in the monkey visual system showed facial motion selectivity. (See Saleem and Logothetis coordinates in Table 1). In comparison, the contrast of fMRI response to motion caused by objects (dynamic objects > static objects) relative to motion caused by faces (dynamic faces > static faces) ($p < 0.0001$, FDR corrected) identified regions mostly in the posterior and middle portions of STS in all 4 monkeys (see Fig. 2B).

3.1.2. Humans—Fig. 3A shows the facial motion selective regions for the representative single human subject, localized by contrasting the fMRI response to motion caused by faces (dynamic faces > static faces) relative to the fMRI responses to motion caused by objects (dynamic objects > static objects) ($p < 0.001$, FDR corrected), as well as from the group results ($p < 0.001$, FDR corrected). The facial motion selective regions were found in three separate foci along the right STS: anterior STS (aSTS), middle STS (mSTS) and posterior STS (pSTS) in all 16 subjects (See Talairach coordinates of averaged activation peaks in Table 2). We did not find any significant areas selective for facial motion in the ventral temporal cortex outside the STS. In comparison, the contrast of fMRI response to motion caused by objects (dynamic objects > static objects) relative to motion caused by faces (dynamic faces > static faces) (individual maps: $p < 0.0001$, FDR corrected; group maps: $p < 0.001$, FDR corrected) identified regions outside of STS (see Fig. 3B).

3.2. Face selectivity

3.2.1. Monkeys—Supplementary Figure S4A shows a map of face-selective regions (face patches) identified by contrasting the fMRI response to static faces with that to static objects ($p < 0.001$, FDR corrected) for each of the four monkeys. Face patches were found bilaterally in the anterior lateral (AL), anterior fundus (AF), middle lateral (ML), middle fundus (MF) of STS in all four monkeys. The location of these face-selective regions is consistent with previous reports in macaque monkeys showing face patches along the superior temporal sulcus within the temporal lobe (Tsao et al., 2008; also see Supplementary Table S1). We did not find significant activation in the posterior occipitotemporal cortex using our threshold of $p < 0.001$ (FDR corrected), indicating that the posterior lateral (PL) face patches were not localized in our study. Activations in the anterior medial (AM) patch within anterior cytoarchitectonic area TE were found bilaterally in monkey F and in the right hemisphere of monkey K. The monkey amygdala is considered to be a face-responsive region since it typically does not show a significantly greater response to static faces than to static objects, but rather shows a significantly greater response to static faces than static scrambled faces (Hadj-Bouziane et al., 2008). The right column of Supplementary Figure S4A shows that face-responsive regions were identified bilaterally in the dorsal portion of

the basal and lateral nuclei of the amygdala in all four monkeys, consistent with previous studies (Hadj-Bouziane et al., 2008, 2012).

Supplementary Figure S4B shows the dynamic face-selective regions identified by contrasting the fMRI response to dynamic faces with that to dynamic objects ($p < 0.001$, FDR corrected) for each of the four monkeys. This dynamic face localizer map essentially reproduced the known face-selective regions, with significant activations found in AL, AF, ML, and MF bilaterally in all four monkeys, and significant activation in the AM face patch bilaterally in monkey F. For the significantly activated face patches of the four monkeys, the activation peak in the dynamic and static localizers were at the same brain location (see Supplementary Table S1 for the activation peak coordinates and Supplementary Table S2 for the distance between the activation peaks for static and dynamic face selective regions).

By examining the maps of facial motion selectivity and face selectivity with a conjunction analysis, we found that the areas of overlap for the two contrast maps were located in the anterior fundus of STS for monkeys F, R and K bilaterally, and for monkey B in the right hemisphere. Further examination revealed that the facial motion selective activation peaks in anterior STS and face selective peaks in AF face patches showed good overlap in all four monkeys (see Supplementary Table S2 for distance between the activation peaks for static and facial motion selective regions), suggesting that facial motion face patches and face patches in AF share the same neural substrate.

3.2.2. Humans—Supplementary Figure S5A shows face-selective regions in the right and left hemisphere for a representative human subject localized by contrasting the fMRI response to static faces with that to static objects ($p < 0.001$, FDR corrected). Regions selectively activated by faces were consistently found in the right hemisphere of inferior lateral occipital gyrus (occipital face area, OFA), lateral fusiform gyrus (fusiform face area, FFA), posterior superior temporal sulcus (pSTS), anterior inferior temporal cortex (aIT) and the dorsolateral portion of the amygdala, in all 16 subjects (see Supplementary Table S3). This is consistent with previous studies that have localized face-selective regions in humans (e.g., Kanwisher et al., 1997; McCarthy et al., 1997; Harris et al., 2012; Axelrod and Yovel, 2013). Additionally, significant face selectivity was found in right mid-STS (mSTS) in 5 of the 16 subjects ($p < 0.001$, FDR corrected).

Supplementary Figure S5B shows the dynamic face-selective regions identified by contrasting the fMRI response to dynamic faces with that to dynamic objects ($p < 0.001$, FDR corrected) for a representative subject. Dynamic face selectivity was found not only in the known face-selective FFA, OFA, aIT, posterior STS, and the amygdala, but also in the middle and anterior STS. The dynamic face-selective region was found in the right posterior STS of all 16 subjects, in the right middle STS of 14 subjects, and in the right anterior STS of 8 subjects ($p < 0.001$, FDR corrected) (see Supplementary Table S3). For those subjects showing significantly activated face-selective regions in both the dynamic and static localizers, their Talairach coordinates of activation peaks were very close (see Supplementary Table S3 for the activation peak coordinates and Supplementary Table S4 for the distance between the activation peaks for static and dynamic face selective regions).

By examining the human maps of facial motion selectivity and face selectivity for each individual subject with a conjunction analysis, we found that the areas of overlap were located in the right posterior STS in all 16 subjects, and in the right middle STS in 5 subjects. Further examination revealed that the Talairach coordinates of activation peaks for facial motion selective and face selective areas in both the posterior STS and middle STS were close (see Supplementary Table S4 for the distance between the activation peaks for static and facial motion selective regions), suggesting good overlap in these areas. Critically however, we found no overlap area in the right anterior STS in the conjunction analysis, indicating that facial motion selective regions in anterior STS were not face-selective.

3.3. Region of interest (ROI) definitions

3.3.1. Monkeys—Our monkey ROIs included the following: AL, AF, ML, MF and the amygdala, in both left and right hemispheres. The ROI in AF, AL, ML, and MF were selected from each monkey's static face selectivity map, by drawing a sphere of 6-mm radius around the peak of the activation ($p < 0.001$, FDR corrected), excluding the voxels that did not meet the significant criterion. The ROI in the amygdala was selected by manually drawing a region within the amygdala's anatomical boundaries, around the peak of the face-responsive activation (static face > static scrambled face; $p < 0.001$, FDR corrected). Because AM was only identified in a small proportion of monkeys in the face-selectivity maps and did not show any activation in the facial motion selectivity maps, we did not include AM face patches in the subsequent ROI analysis. The Saleem and Logothetis coordinates of the activation peaks for these ROIs are shown in Supplementary Table S1.

3.3.2. Humans—Our human ROIs included the following regions: OFA, FFA, aIT, pSTS, mSTS, aSTS and the amygdala. The ROIs in OFA, FFA and aIT were selected from each individual subject's face selectivity map, and the ROIs in pSTS, mSTS, and aSTS were selected from each individual subject's facial motion selectivity map, each by drawing a sphere of 6-mm radius around the peak of the activation ($p < 0.001$, FDR corrected), excluding the voxels that did not meet the significance criterion. The ROI for the amygdala was selected by manually drawing a region around the peak of the face-selective activation ($p < 0.001$, FDR corrected) that fell within the amygdala's anatomical boundaries. Please note that our human ROIs in the right hemisphere were found in all 16 subjects. In the left hemisphere, the ROI in pSTS was found in all 16 subjects, the ROI in mSTS was found in 11 subjects, and the ROI in aSTS was found in 13 subjects ($p < 0.001$, FDR corrected). The Talairach coordinates of the activation peaks for these ROIs are shown in Supplementary Table S3.

3.4. ROI analysis of PSC data

3.4.1. Monkeys—In monkeys, the fMRI response amplitudes (averaged across left and right hemispheres) evoked by each stimulus condition in each of the 5 ROIs are shown in Fig. 4A–E. The PSC data for dynamic monkey faces, dynamic objects, static monkey faces and static objects from each run were averaged within each ROI and submitted to a category (levels: faces, objects) x motion (levels: dynamic, static) x hemisphere (levels: left, right) x ROIs repeated-measures analysis of variance (ANOVA). To understand the nature of the interactions, we also performed a category (levels: face, object) x motion (levels: dynamic,

static) x hemisphere (levels: left, right) repeated-measures ANOVA for each ROI. For all of the ANOVAs, we included the monkey (levels: Monkey F, Monkey R, Monkey K) as a between-subject factor. We were especially interested in the interaction between category and motion because it, to a certain extent, also reflects the selectivity of facial motion in the ROI (Pitcher et al., 2011; Polosecki et al., 2013), which would verify the facial motion selective areas that we identified from the facial motion selectivity maps. Post-hoc analyses were used to identify specific effects, and p values were Bonferroni corrected for the number of pair-wise comparisons.

The repeated measures ANOVA of the PSC data (averaged across left and right hemispheres) revealed significant main effects of category ($F_{(1,32)} = 308.24$; $p < 10^{-17}$), motion ($F_{(1,32)} = 140.65$; $p < 10^{-12}$) and ROI ($F_{(4,128)} = 148.35$; $p < 10^{-46}$), but not of hemisphere ($F_{(1,32)} = 1.67$; $p = 0.21$). Significant 2-way interactions were found between category and ROI ($F_{(4,128)} = 124.38$; $p < 10^{-42}$), and motion and ROI ($F_{(4,128)} = 54.57$; $p < 10^{-26}$), and a significant 3-way interaction was found between category, motion and ROI ($F_{(4,128)} = 7.98$; $p < 10^{-5}$). The 4-way category x motion x ROI x hemisphere interaction was not significant ($F_{(1,128)} = 0.18$; $p = 0.95$; also see Supplementary Table S5 for all statistical values). These results indicate that different ROIs responded differently to face and object stimuli and that these responses were modulated by stimulus motion.

By performing a separate repeated-measures ANOVA for each ROI, we found a significant main effect of category in each ROI (AF: $F_{(1,32)} = 74.55$, $p < 10^{-9}$; MF: $F_{(1,32)} = 186.66$, $p < 10^{-14}$; AL: $F_{(1,32)} = 392.05$, $p < 10^{-18}$; ML: $F_{(1,32)} = 274.50$, $p = 10^{-16}$; amygdala: $F_{(1,32)} = 12.71$, $p = 0.0012$), with the fMRI response to faces significantly greater than that to objects (face > object), confirming a face-selective response in these regions. We also found a significant main effect of motion in all ROIs, except for the amygdala (AF: $F_{(1,32)} = 43.03$, $p < 10^{-6}$; MF: $F_{(1,32)} = 236.42$, $p < 10^{-15}$; AL: $F_{(1,32)} = 62.64$, $p < 10^{-8}$; ML: $F_{(1,32)} = 118.33$, $p < 10^{-11}$; amygdala: $F_{(1,32)} = 2.32$, $p = 0.14$), with the fMRI responses to dynamic stimuli significantly greater than that to static stimuli. Further, a significant interaction was found between motion and category in face patch AF ($F_{(1,32)} = 25.26$; $p < 10^{-4}$) but not in other ROIs (MF: $F_{(1,32)} = 0.025$; $p = 0.87$; AL: $F_{(1,32)} = 0.32$; $p = 0.57$; ML: $F_{(1,32)} = 3.75$; $p = 0.062$; amygdala: $F_{(1,32)} = 0.48$; $p = 0.49$). Bonferroni corrected post-hoc comparisons on AF revealed that the fMRI response to dynamic faces was significantly greater than that to dynamic objects ($p < 10^{-8}$), and the fMRI response to dynamic faces was significantly greater than that to static faces ($p < 10^{-7}$), demonstrating the specialization of AF for facial motion (see Supplementary Table S6).

Fig. 4F shows the facial motion and object motion sensitivity values for each monkey ROI. The PSC values of facial motion and object motion sensitivity were entered into a repeated-measures ANOVA with category (level: face, object), hemisphere (levels: left, right) and ROI as within-subject factors, and monkey (level: Monkey F, Monkey R, Monkey K) as a fixed-effect between-subject factor. The result revealed a significant main effect of ROI ($F_{(4,128)} = 75.70$; $p < 10^{-28}$), but no significant main effect of category ($F_{(1,32)} = 0.79$; $p = 0.38$) or hemisphere ($F_{(1,32)} = 0.083$; $p = 0.78$). Importantly, a significant 2-way interaction between category and ROI ($F_{(4,128)} = 9.76$; $p < 10^{-6}$) was found, indicating that the various ROIs differed in their motion sensitivity to faces and objects. Bonferroni corrected post-hoc

comparisons revealed that only AF showed a significantly greater facial motion sensitivity than object motion sensitivity (AF: $p < 10^{-7}$; MF: $p = 0.89$; AL: $p = 0.59$; ML: $p = 0.33$; amygdala: $p = 0.35$).

3.4.2. Humans—In humans, the fMRI response amplitudes for each stimulus condition in each of the 7 right hemisphere ROIs are shown in Fig. 5A–G. The PSC data for dynamic human faces, dynamic objects, static human faces and static objects for each subject were averaged within each ROI and submitted to a repeated-measures ANOVA with category (levels: face, object), motion (levels: dynamic, static) and ROI as within-subject factors. Similar to the monkey data, we also performed a repeated-measures ANOVA for each ROI, with category (levels: face, object) and motion (levels: dynamic, static) as within-subject factors.

The repeated-measures ANOVA of the PSC data revealed significant main effects of category ($F_{(1,15)} = 403.11$; $p < 10^{-11}$), motion ($F_{(1,15)} = 63.40$; $p < 10^{-6}$) and ROI ($F_{(6,90)} = 44.36$; $p < 10^{-24}$). In addition, there were significant 2-way interactions between category and motion ($F_{(1,15)} = 21.83$; $p = 0.00030$), between motion and ROI ($F_{(6,90)} = 16.47$; $p < 10^{-11}$), and between category and ROI ($F_{(6,90)} = 162.83$; $p < 10^{-45}$). The 3-way interaction between category, motion and ROI was also significant ($F_{(6,90)} = 28.74$; $p < 10^{-18}$; See also Supplementary Table S7).

A repeated-measures ANOVA of the PSC data for each ROI revealed a significant main effect of category in all right hemisphere ROIs (FFA: $F_{(1,15)} = 392.63$; $p < 10^{-11}$; OFA: $F_{(1,15)} = 202.92$; $p < 10^{-9}$; aIT: $F_{(1,15)} = 295.73$; $p < 10^{-10}$; aSTS: $F_{(1,15)} = 52.73$; $p < 10^{-5}$; mSTS: $F_{(1,15)} = 56.30$; $p < 10^{-5}$; pSTS: $F_{(1,15)} = 219.10$; $p < 10^{-9}$; amygdala: $F_{(1,15)} = 112.54$; $p < 10^{-7}$), with the fMRI response to faces significantly greater than that to objects, confirming the face selectivity in these regions. We also found a significant main effect of motion in all ROIs, except for the amygdala (FFA: $F_{(1,15)} = 41.81$; $p < 10^{-4}$; OFA: $F_{(1,15)} = 56.96$; $p < 10^{-5}$; aIT: $F_{(1,15)} = 47.35$; $p < 10^{-5}$; aSTS: $F_{(1,15)} = 32.99$; $p < 10^{-4}$; mSTS: $F_{(1,15)} = 36.96$; $p < 10^{-4}$; pSTS: $F_{(1,15)} = 101.73$; $p < 10^{-7}$; amygdala: $F_{(1,15)} = 2.29$; $p = 0.15$), with the fMRI response to dynamic stimuli significantly greater than that to static stimuli, thus confirming the significant effect of motion in our ROIs (except for the amygdala). More importantly, we found a significant interaction between motion and category in the aSTS ($F_{(1,15)} = 87.45$; $p < 10^{-6}$), mSTS($F_{(1,15)} = 83.77$; $p < 10^{-6}$) and pSTS ($F_{(1,15)} = 12.11$; $p = 0.0034$), but not in the other ROIs (aIT: $F_{(1,15)} = 0.04$; $p = 0.85$; FFA: $F_{(1,15)} = 0.06$; $p = 0.81$; OFA: $F_{(1,15)} = 3.70$; $p = 0.074$; amygdala: $F_{(1,15)} = 2.82$; $p = 0.11$). Bonferroni corrected post-hoc comparisons showed that, in all three regions of the STS, the fMRI response to dynamic faces was significantly greater than that to dynamic objects (aSTS: $p < 10^{-8}$; mSTS: $p < 10^{-7}$; pSTS: $p < 10^{-9}$) and the fMRI response to dynamic faces was also significantly greater than to that to static faces (aSTS: $p < 10^{-6}$; mSTS: $p < 10^{-6}$; pSTS: $p < 10^{-8}$). These results demonstrated the specialization of aSTS, mSTS, and pSTS for facial motion (See Supplementary Table S8).

Fig. 5H shows the facial motion and object motion sensitivity for each ROI in humans. A repeated-measures ANOVA revealed a significant main effect of category ($F_{(1,15)} = 21.83$; $p = 0.00030$) and of ROI ($F_{(6,90)} = 16.47$; $p < 10^{-11}$), as well as a significant interaction

between category and ROI ($F_{(6,90)} = 28.73$; $p < 10^{-18}$), indicating that the motion preference of the various ROIs differed between faces and objects. Bonferroni corrected post-hoc comparisons revealed that aSTS, mSTS and pSTS showed a significantly greater preference for facial motion than object motion (aSTS: $p < 10^{-6}$; mSTS: $p < 10^{-6}$; pSTS: $p = 0.003$). In summary, our results indicated that all three regions in STS, aSTS, mSTS and pSTS, showed consistent facial motion preference.

3.5. ROI analysis of facial-motion selectivity

3.5.1. Monkeys—Fig. 6A shows the facial motion selectivity [(dynamic faces - static faces) - (dynamic non-face objects - static non-face objects)] for each monkey ROI. To understand the pattern of facial motion selectivity among these ROIs, the selectivity values were entered into a repeated-measures ANOVA with hemisphere (levels: left, right) and ROI as within-subject factors, and monkey (level: Monkey F, Monkey R, Monkey K) as a between-subject factor. The result revealed a significant main effect of ROI ($F_{(4,128)} = 8.22$; $p = 0.000006$), but not of hemisphere ($F_{(1,32)} = 0.18$; $p = 0.68$). The 2-way interaction between ROI and hemisphere ($F_{(4,128)} = 0.64$; $p = 0.63$) was not significant. These indicated that different ROIs had different facial motion selectivity, while facial motion selectivity did not show hemisphere bias. Bonferroni corrected post-hoc comparisons revealed that AF showed significantly greater facial motion selectivity than MF ($p = 0.00042$), ML ($p < 10^{-6}$), and amygdala ($p = 0.00099$), and a trend for greater facial motion selectivity than AL ($p = 0.087$) - no other pair of ROIs showed a significant facial motion selectivity difference. In addition, a one-sample *t*-test revealed that only AF showed significant non-zero facial motion selectivity ($t_{(34)} = 7.15$; Bonferroni corrected $p = 1.43 \times 10^{-7}$, correction based on performing a total of 5 one-sample t-tests for the 5 ROIs). In summary, these results indicated that only AF, and none of the other brain regions, showed facial motion specialization, and AF was the most prominent region, relative to all other brain regions, to show significant facial motion selectivity.

3.5.2. Humans—Fig. 6B shows the facial motion selectivity values for each human right hemisphere ROI. A one-way repeated-measures ANOVA of these values, with ROI as within-subject factor, revealed a significant main effect of ROI ($F_{(6,90)} = 28.73$; $p < 10^{-18}$). Bonferroni corrected post-hoc comparisons revealed that aSTS showed significantly greater facial motion selectivity than pSTS ($p = 0.0096$), aIT ($p = 0.000054$), FFA ($p = 0.00017$), OFA ($p = 0.000052$) and amygdala ($p < 10^{-6}$); mSTS showed significantly greater facial motion selectivity compared to aIT ($p = 0.000062$), FFA ($p = 0.00034$), OFA ($p = 0.00016$), and amygdala ($p < 10^{-6}$); pSTS showed a significantly greater facial motion selectivity than FFA ($p = 0.001$), OFA ($p = 0.035$) and amygdala ($p = 0.00013$). In addition, a one-sample *t*-test revealed that aSTS and mSTS showed significant non-zero facial motion selectivity (aSTS: $t_{(15)} = 9.35$, $p = 8.39 \times 10^{-7}$; mSTS: $t_{(15)} = 9.15$, $p = 1.11 \times 10^{-6}$; Bonferroni corrected based on 7 one-sample t-tests, one for each ROI); pSTS showed a trend for significant non-zero facial motion selectivity ($t_{(15)} = 3.48$, $p = 0.024$; Bonferroni corrected); aIT, FFA, OFA and amygdala did not show significant facial motion selectivity (aIT: $t_{(15)} = 0.19$, $p > 0.99$; FFA: $t_{(15)} = 0.25$, $p > 0.99$; OFA: $t_{(15)} = -1.92$, $p = 0.52$; amygdala: $t_{(15)} = -1.68$, $p = 0.80$; Bonferroni corrected). Our ROI results in the right hemisphere indicated that only aSTS and mSTS regions showed consistent facial motion selectivity ($p < 0.01$), and

the facial motion preference in aSTS was the most prominent among all visual regions of the human brain.

3.5.3. Left hemisphere ROIs—Although not all subjects showed significant clusters in the left hemisphere for each ROI, we examined facial motion selectivity in the left hemisphere for the subset of subjects that did (see Supplementary Table S3 for number of subjects used for each ROI). Supplementary Figure S6 shows the fMRI response amplitudes for each stimulus condition in each of the 7 left hemisphere ROIs. A category (levels: face, object) x motion (levels: dynamic, static) two-way repeated-measures ANOVA of the PSC data for each ROI revealed a significant interaction between motion and category in the aSTS ($F_{(1,12)} = 91.29$; $p < 10^{-6}$), mSTS ($F_{(1,10)} = 25.70$; $p = 0.00049$) and pSTS ($F_{(1,15)} = 9.28$; $p = 0.0082$), but not in the other ROIs (aIT: $F_{(1,11)} = 0.064$; $p = 0.81$; FFA: $F_{(1,15)} = 0.22$; $p = 0.65$; OFA: $F_{(1,13)} = 1.74$; $p = 0.21$; amygdala: $F_{(1,14)} = 3.69$; $p = 0.075$) (see Supplementary Table S9). Bonferroni corrected post-hoc comparisons showed that, in all three regions of the STS, the fMRI response to dynamic faces was significantly greater than that to dynamic objects (aSTS: $p < 10^{-7}$; mSTS: $p = 0.00032$; pSTS: $p < 10^{-8}$) and the fMRI response to dynamic faces was also significantly greater than to that to static faces (aSTS: $p < 10^{-6}$; mSTS: $p = 0.000086$; pSTS: $p < 10^{-7}$). These results demonstrated the sensitivity of aSTS, mSTS, and pSTS in the left hemisphere for facial motion.

Supplementary Figure S6H shows the facial motion and object motion sensitivity for each left hemisphere ROI. A two-way repeated-measures ANOVA of the sensitivity data for 11 subjects (for whom significant activation clusters were seen in all 7 ROIs in the left hemisphere) revealed a significant main effect of category ($F_{(1,10)} = 11.70$, $p = 0.0065$) and of ROI ($F_{(6,60)} = 3.66$, $p = 0.0037$), as well as a significant interaction between category and ROI ($F_{(6,60)} = 10.37$, $p < 10^{-7}$), indicating that the motion preference of the various ROIs differed between faces and objects. Bonferroni corrected post-hoc comparisons revealed that aSTS and mSTS showed a significantly greater preference for facial motion than object motion (aSTS: $p < 10^{-5}$; mSTS: $p = 0.00049$), while pSTS, aIT, FFA, OFA and amygdala did not show a significant preference for facial motion than object motion (pSTS: $p = 0.059$; aIT: $p = 0.65$; FFA: $p = 0.63$; OFA: $p = 0.21$; amygdala: $p = 0.087$). A one-way repeated-measures ANOVA (with ROI as within-subject factor) of facial motion selectivity values for 11 subjects, revealed a significant main effect of ROI ($F_{(6,60)} = 10.37$; $p < 10^{-7}$). Bonferroni corrected post-hoc comparisons revealed that aSTS showed significantly greater facial motion selectivity than aIT ($p = 0.0037$), FFA ($p = 0.0010$), OFA ($p = 0.00019$) and amygdala ($p = 0.00093$), and a trend for greater facial motion selectivity compared to pSTS ($p = 0.027$); mSTS showed significantly greater facial motion selectivity compared to amygdala ($p = 0.00078$). Further, one-sample t-tests revealed that aSTS and mSTS showed significant non-zero facial motion selectivity (aSTS: $t_{(12)} = 11.46$, $p = 5.66 \times 10^{-7}$; mSTS: $t_{(10)} = 5.07$, $p = 0.0034$; Bonferroni corrected based on 7 one-sample t-tests, one for each ROI); pSTS, aIT, FFA, OFA and amygdala did not show significant facial motion selectivity (pSTS: $t_{(15)} = 3.05$, $p = 0.057$; aIT: $t_{(11)} = 0.25$ $p = 0.99$; FFA: $t_{(15)} = 0.46$, $p > 0.99$; OFA: $t_{(13)} = -1.32$, $p > 0.99$; amygdala: $t_{(14)} = -1.92$, $p = 0.53$; Bonferroni corrected). Thus, results for left hemisphere ROIs from a subset of subjects were similar to those from right hemisphere ROIs, and supported the conclusion that bilateral aSTS and mSTS regions

showed strong and consistent facial motion selectivity ($p < 0.01$) and that the facial motion preference in aSTS was the most prominent among all visual regions of the human brain.

4. Discussion

In the current study, we used fMRI to investigate the extent to which the specialization for facial motion is represented in the visual system of monkeys and humans, and if they share similar neural substrates. We explicitly defined the areas specialized for facial motion as those significantly more active when subjects perceive non-rigid motion in faces (dynamic faces > static faces) than when they perceive the non-rigid motion in non-face objects (dynamic non-face objects > static non-face objects). Taking advantage of equivalent motion energies across the dynamic conditions in our study, we were able to map the facial motion selectivity in the visual system of both monkeys (at 4.7T) and humans (at 7T). Our results showed that, in all four monkeys scanned, significant activations evoked by non-rigid facial motion were found in the fundus of anterior STS, within anterior inferior temporal cortex of cytoarchitectonic area TE. This facial motion region overlapped the anterior fundus face-selective patch AF. In humans, facial motion activated three separate foci in the right STS: anterior STS (aSTS), middle STS (mSTS) and posterior STS (pSTS), with the aSTS showing the greatest selectivity for facial motion. Taken together, our results indicate that monkeys and humans share similar neural substrates within the anterior STS for the processing of facial motion.

4.1. Facial motion selectivity in monkeys and humans

Our current study explicitly defines facial motion selectivity in both monkeys and humans, and is the first to map facial motion selectivity at the whole brain level in the two species.

4.1.1. Monkeys—In monkeys, by contrasting the fMRI response to motion caused by monkey faces to that caused by non-face objects, we accurately mapped the facial motion selectivity in the fundus of STS that overlapped the anterior fundus (AF) face patches bilaterally. No other brain region in the monkey visual system showed facial motion selectivity. In the ROI analysis, we found a significant interaction between ROI and motion category preference (the sensitivity of facial motion and object motion), with AF showing a significantly greater facial motion preference than object motion preference. Further ROI selectivity analysis confirmed that only AF, and none of the other brain regions, showed specialization for facial motion. Thus, both our whole brain mapping and ROI analysis showed that only AF is specialized for facial motion.

Previous studies in monkeys have suggested that specialized neural substrates for processing facial motion exist in STS, especially in anterior STS, but results from those studies are mixed (Furl et al., 2012; Polosecki et al., 2013; Fisher and Freiwald, 2015a). Polosecki et al. (2013) reported that facial motion is not represented in the fundus of STS but rather within the anterior lateral (AL) and anterior medial (AM) face patches, while Fisher and Freiwald (2015a) reported facial motion selectivity in all STS face patches, with a preference for facial motion most pronounced in patches along the fundus and dorsal bank of the STS. Additionally, both of these studies used an ROI approach, limiting their investigation of facial motion selectivity to the pre-defined face patches.

Our results are inconsistent with those of both Polosecki and colleagues (Polosecki et al., 2013), who reported that AF is not specialized for facial motion, and of Fisher and Freiwald (2015a), who reported that all face patches, including AF, are specialized for facial motion. The discrepancy between our ROI results and the two previous studies may be due to several reasons. First, both Fisher and Freiwald (2015a) and Polosecki and colleagues (2013) used dynamic monkey faces that portrayed facial expressions, while our study used only neutral faces. Further, the monkey face videos in Fisher and Freiwald's study also contained rigid head motion, while our study used only non-rigid facial motion. Therefore, it is possible that differences in the stimuli may be the reason for the discrepancy in the results. Another possible reason for the discrepancy is that we used a different method to control motion energy in the stimuli. Previous studies (Fisher and Freiwald, 2015a; Polosecki et al., 2013) did not equate the motion energy contained in the dynamic face and object stimuli; instead they used the fMRI response in motion-sensitive areas to infer that the motion energy was not greater in the dynamic face stimuli than the dynamic object stimuli. This method of evaluating motion energy can be inaccurate because neural responses in motion-sensitive areas are modulated not only by motion energy, but also by stimulus type. In fact, with motion energy equated between face and object stimuli in our study, a similar analysis showed greater object motion preference in MST/FST (termed 'general motion selective area' in Fisher and Freiwald, 2015a; Dubner and Zeki, 1971; Desimone and Ungerleider, 1986). This suggests a greater sensitivity to object motion in these regions, and thus, similar fMRI responses to object and facial motion cannot be used as support for equal motion energy in underlying stimuli.

In addition to facial motion, macaque AF has also been reported to be involved in aspects of social cognition. In a recent single-unit recording study (McMahon et al., 2015), neurons in macaque AF were shown to be sensitive to the social content of faces, while macaques freely viewed complex videos rich with natural social content. In addition, in a recent fMRI study (Fisher and Freiwald, 2015b), the macaque AF face patch exhibited a super-additive neural response to face and body integration, suggesting a whole-agent selectivity of AF in social cognition. Because facial motion sensitivity, found here in macaque AF, conveys emotional expressions, our results provide additional evidence that AF may be a key region for integrating face information within a social context.

4.1.2. Humans—In our human subjects, by contrasting the fMRI response to motion caused by human faces to that caused by non-face objects, we consistently identified three separate foci of activation in the right hemisphere (and to some extent in the left hemisphere). These regions were all located within the STS, with activated foci in anterior, middle and posterior regions. Our ROI analysis showed a significant interaction between ROI and motion category preference (i.e., sensitivity to facial motion and object motion), with only aSTS and mSTS showing a significantly greater preference for facial motion than object motion. Further, our ROI selectivity analysis confirmed that only aSTS and mSTS, and none of the other brain regions, showed consistent facial motion selectivity.

Evidence from previous studies in humans also suggests that specialized neural circuitry for processing facial motion exists in STS (Fox et al., 2009; Pitcher et al., 2011; Schultz et al., 2013). The two previous studies examining interactions between category and motion

revealed significant facial motion selectivity in face-selective regions of STS. However, both these studies examined facial motion selectivity only in predefined face-selective ROIs (Pitcher et al., 2011; Polosecki et al., 2013), and it remained unclear how facial motion selectivity is distributed outside of these ROIs across the whole visual system.

Our results are consistent with the previous two studies (Pitcher et al., 2011; Polosecki et al., 2013) reporting the involvement of face-selective regions within STS for facial motion selectivity. However, instead of examining facial motion selectivity in predefined face-selective ROIs, our facial motion selectivity map revealed more activated foci along STS than previous studies. Moreover, we found in our motion-sensitivity analysis that the anterior STS facial motion selective region showed significantly greater facial motion preference than middle STS or posterior STS regions, suggesting that human anterior STS is preferentially sensitive to facial motion relative to more posterior STS regions.

Anterior STS is typically not thought of as a face-selective region because it is not usually activated more by static faces than static objects in traditional face localizers (Pitcher et al., 2011). Even a contrast of dynamic faces vs. dynamic non-face objects does not reliably activate this region and, if activated, the size of the activated region is usually relatively small (Fox et al., 2009; Pitcher et al., 2011). Anterior STS is also not a frequently reported region for motion. In many previous studies not using face stimuli, the contrast of dynamic stimuli with static stimuli has not evoked significant activation in anterior STS (Polosecki et al., 2013). However, in our current study, by mapping facial motion selectivity contrasting the fMRI response to motion caused by faces (dynamic faces > static faces) relative to the fMRI responses to motion caused by objects (dynamic objects > static objects), we were able to reliably find significant activation in the anterior STS, confirming facial motion selectivity in this region. We also suggest that, with motion energy equivalent across the dynamic stimuli, the contrast of facial motion with that to object motion can be a reliable way to identify facial motion selective areas.

In addition to anterior STS, we found a region in the middle portion of STS that showed significant facial motion selectivity. The averaged Talairach coordinate of the peak voxels for this region was [48, −13, −8], which is located between facial motion regions in the anterior STS (Talairach coordinate [57, 8, −16]) and in the posterior STS (Talairach coordinate [52, −38, 2]). The region in the middle STS partially overlapped face-selective regions in the middle and anterior STS previously identified by contrasting activation to dynamic faces with that to dynamic objects (Fox et al., 2009; Pitcher et al., 2011). In our study, this middle STS region was distinct from both the anterior and posterior STS regions in the facial motion map in all 16 subjects, confirming the middle STS as a discrete region for processing facial motion. Subsequent ROI analysis showed a significantly greater preference for facial motion than object motion, and significant facial motion selectivity in the middle STS, confirming the specialization of this region for facial motion.

As a human face-selective region, the posterior STS is often presumed to process the dynamic aspect of faces (Bruce and Young, 1990; Haxby et al., 2000). Previous neuroimaging studies have reported facial motion specialization in the posterior STS, particularly in the right hemisphere (Puce et al., 1998; Pitcher et al., 2011; Polosecki et al.,

2013). Consistent with these findings, our whole-brain analysis identified regions along the STS that are specialized for facial motion, including the posterior STS, and our ANOVA of PSC data found significant interactions between category (levels: faces, objects) and motion (levels: dynamic, static) in posterior STS (left: $p = 0.0082$; right: $p = 0.0034$). However, this region showed very weak facial motion selectivity (left pSTS: $p = 0.057$; right pSTS: $p = 0.024$), suggesting that the posterior STS is not a reliable region for facial motion selectivity.

The posterior STS has previously been reported to be activated not just by facial motion but by a variety of non-face motion stimuli, such as the action of walking, hand grasping, and point-light biological motion (Puce et al., 1998; Beauchamp, 2002, 2003; see Bernstein and Yovel, 2015; Duchaine and Yovel, 2015 for review). In our current study, the posterior STS did not exhibit significant facial motion selectivity, which differs from the anterior STS, which showed strong facial motion selectivity (Fig. 6). Considering that the facial motion preference in anterior STS was significantly greater than in posterior STS, our results suggest that the specialization for processing of facial motion information increases as one progresses along the STS from posterior to anterior. Thus, pSTS is involved in generalized processing of motion (facial and object), and likely transmits this information anteriorly to middle and anterior STS regions, which in turn selectively extract information related to facial dynamics. Further, These results are consistent with posterior to anterior STS gradients reported in other modalities such as voice identity recognition (Deen et al., 2015; Schall et al., 2015) and support the notion that the STS in general is involved in multisensory integration, and the posterior STS in particular may be involved in cross-modal binding between auditory and visual stimuli (Beauchamp et al., 2004, 2015).

It is widely acknowledged that ventral stream areas FFA and OFA are involved in face processing, in that they are activated more by faces than by complex non-face objects (Kanwisher and Yovel, 2006). The two previous studies probing facial motion selectivity (Pitcher et al., 2011; Polosecki et al., 2013) did not find a significant category by motion interaction in the FFA and OFA, indicating that these two regions do not show facial motion selectivity. In both our whole brain mapping and ROI analyses, we too did not find significant facial motion selectivity in either the FFA or OFA; indeed no other region in the ventral stream (e.g. aIT) showed facial motion selectivity, consistent with previous reports of a lack of facial motion specialization in the human ventral stream (Pitcher et al., 2011; Polosecki et al., 2013).

Overall, our findings indicate that the selectivity for facial motion in humans is represented in regionally specific brain regions along the STS, with the preference increasing as one moves anteriorly.

4.2. Face selectivity in monkeys and humans

Studies of functional specialization typically locate regions in the brain that are domain-specific for certain categories of stimuli. The most well-known example is the localization of face-selective regions in human and nonhuman primates (Kanwisher et al., 1997; Tsao et al., 2008), which have been delineated by contrasting the fMRI response evoked by images of faces (most often static) with that evoked by images of non-face objects, typically in a block design.

4.2.1. Monkeys—In monkeys, by contrasting the fMRI response evoked by images of static monkey faces with that evoked by images of static non-face objects, we identified bilateral face patches in anterior lateral (AL), anterior fundus (AF), middle lateral (ML), and middle fundus (MF) STS and patches in the anterior medial (AM) temporal lobe of four macaques, consistent with previous reports of face selectivity in the monkey brain (e.g., Tsao et al., 2008; Hadj-Bouziane et al., 2008).

The contrast of the fMRI response to dynamic faces with that to dynamic objects evoked significant activations in AL, AF, ML, and MF face patches of STS in all four monkeys, matching our findings with static face patches. However, we also found one monkey that did not show an AL dynamic face patch in the left hemisphere, two monkeys with AL dynamic face patches in the right hemisphere that were smaller than their AL static face patches, and one monkey that showed a static AM face patch but did not show a dynamic AM face patch in the right hemisphere. Our dynamic face localizer results are similar to those of Polosecki and colleagues (Polosecki et al., 2013), who reported that the dynamic face localizer reproduced the known static face patches, but differ from those of Fisher and Freiwald's (2015a), who reported that the dynamic face localizer not only activated all the known static face patches, but also activated a new region, termed the middle dorsal face patch (MD), located on the anterodorsal bank of the STS.

In our monkeys, the contrast of the fMRI response to dynamic faces with that to static faces evoked significant activations throughout the fundus of STS, especially within the anterior STS. Two studies have previously reported the mapping of facial motion sensitivity by contrasting the fMRI response to dynamic faces with that to static faces (Furl et al., 2012; Fisher and Freiwald, 2015a). Our results are largely consistent with those reported in these studies, though also showing small discrepancies: Furl et al. (2012) reported that facial motion sensitivity (dynamic > static faces) evoked in posterior, middle and anterior STS had distinct peak activations from the peaks of the face-selective regions. In contrast, Fisher and Freiwald (2015a) reported that facial motion sensitivity so defined activated a subset of face patches, extending throughout the motion-sensitive areas in STS.

4.2.2. Humans—By contrasting the fMRI response evoked by images of static faces with that evoked by images of static non-face objects in our human subjects, we identified face-selective regions in the lateral fusiform gyrus (FFA), inferior lateral occipital gyrus (OFA), posterior superior temporal sulcus (pSTS), anterior inferior temporal cortex (aIT) and the dorsolateral portion of the amygdala. This is consistent with previous studies that have localized face-selective regions in humans (e.g., Kanwisher et al., 1997; McCarthy et al., 1997; Kriegeskorte et al., 2007; Harris et al., 2012; Axelrod and Yovel, 2013). In our study, the face-selective regions in the right hemisphere of humans were more consistently activated than those in the left hemisphere, which is also consistent with previous studies reporting the lateralization of face selectivity in humans (Kanwisher et al., 1997; Yovel et al., 2008; Willems et al., 2010).

In humans, the contrast of dynamic faces vs. dynamic objects revealed face-selective regions not only in FFA, OFA, aIT, posterior STS, and the amygdala in all 16 subjects, but also in the middle STS in 14 subjects and the anterior STS in 8 subjects. Our results are consistent

with two previous dynamic face localizer studies, including Fox et al. (2009) who reported that the dynamic face localizer activated more regions in the extended system of face perception, including the middle superior temporal sulcus, and Pitcher et al. (2011) who found strong activations to their dynamic localizer in the STS, with a region in the anterior STS responding to dynamic faces only.

The contrasts of dynamic stimuli with static stimuli, dynamic faces with static faces, and dynamic objects with static objects showed similarly activated brain areas along the middle and posterior STS, while the contrast of dynamic faces with static faces additionally showed activation in more anterior STS regions. These results are largely consistent with previous findings (Schultz and Pilz, 2009; Schultz et al., 2013). In their study examining facial motion sensitivity, Schultz and Pilz (2009) found stronger activations in STS and medial temporal gyrus (MTG, including hMT+/V5) in response to dynamic faces compared to static faces.

4.3. Amygdala is not specialized for facial motion in humans or monkeys

We did not find a significantly greater preference for facial motion than object motion, or significant facial motion selectivity in the amygdala in either species, indicating that the human and monkey amygdala is not specialized for facial motion. For motion sensitivity, we did not find any significant activation in the amygdala by contrasting either dynamic faces with static faces or contrasting dynamic non-face objects with static non-face objects in either our human or monkey subjects, indicating that the amygdala is not sensitive to either facial motion or object motion. One previous study reported increased amygdala activation for point-light biological motion (Bonda et al., 1996). However, this finding may have been due to the affective content of the point-light stimuli (figures appeared to be dancing) rather than biological motion per se (Herrington et al., 2011).

4.4. Distinct pathway for facial motion in humans and monkeys

It has been proposed that face processing in humans is transmitted along two distinct neuroanatomical visual pathways: a lateral STS pathway that is presumed to process the changeable aspects of faces, such as emotional expression, and a ventral pathway that is presumed to process the invariant aspects of faces, such as facial identity (Bruce and Young, 1990; Haxby et al., 2000; Puce et al., 1998; Allison et al., 2000; Andrews and Ewbank, 2004). Recently, a review of neuroimaging studies proposed a different neural framework for face processing, with the changeable and invariant facial information replaced by motion and form information, respectively, as the functional division between the lateral and ventral pathways (Bernstein and Yovel, 2015). Thus, in this new framework, human face-selective regions in lateral STS, such as pSTS, are tuned to facial motion, while the human face-selective regions located ventrally, including OFA, FFA and aIT, are tuned to form. Our results are consistent with this framework. We found three separate foci in the human lateral pathway along the right STS that were consistently specialized for facial motion: anterior STS, middle STS and posterior STS, while we did not find any regions specialized for facial motion in ventral cortical visual regions, including OFA, FFA, and aIT. In monkeys, the region that was specialized for facial motion was found in the anterior fundus face patch of STS, while the lateral face patches AL and ML did not show any specialization for facial

motion. Because the dynamic faces in our study were all front-view neutral faces, thereby excluding the effects of emotion and rigid head motion, our results provide strong evidence for the dissociation of neural pathways processing non-rigid facial motion and facial form information in the primate brain of both species. As our main goal in this study was to examine the functional homology of STS regions between the two species, we chose conspecific face stimuli, i.e., different stimuli for monkey and human experiments. We assumed that these different stimulus sets would not evoke different patterns of activations across the two species; however, this should be explicitly examined in a future study by using identical human and monkey face stimulus sets for the two species.

4.5. Increased facial motion specificity along a posterior-to-anterior axis of STS across the two species

Despite many fundamental differences in structural organization between human and macaque brains, the results of our study showed a strikingly similar functional organization for the processing of facial motion across the two species: a facial motion selective area in monkeys is located in the anterior fundus face patch, while a facial motion selective area in humans is most prominent in anterior STS. The facial motion selective area in anterior STS in humans functionally matches the anterior fundus face patch in macaques. In addition, in our monkeys, the whole-brain mapping analysis showed that only the anterior fundus face patch showed facial motion selectivity, and the motion sensitivity analysis showed that facial motion preference in anterior fundus face patch was significantly greater than in the middle fundus face patch, there appears to be an increasing preference for facial motion as one moves anteriorly along the fundus of STS; in our human subjects, because the facial motion preference was most prominent in the anterior STS, and significantly greater there than in posterior STS, there also appears to be a gradient preference for facial motion as one progresses anteriorly along the human STS.

Taken together, our results suggest a homology between human and monkey STS pathways, with similar neural substrates in both species within the anterior STS for the processing of non-rigid facial motion. It should be noted that despite the similarity in facial motion selectivity in anterior STS regions of both species, these regions differed in their degree of face selectivity (monkey AF was strongly face selective, while human aSTS was not). There are also reports of functional distinctions between these two regions (Zhu et al., 2013) in processing dynamic facial expressions, thus the homology between these two regions may not be straightforward. However, it should not be ignored that there are still some functional differences between humans and monkey STS pathways: monkey anterior fundus face patches show consistent face selectivity (contrasting static faces with static object), while human anterior STS does not show face selectivity (using the same contrast). Furthermore, a previous study reported a distinction between human and monkey STS in processing dynamic facial expressions (Zhu et al., 2013), again suggesting a functional difference between STS in the two species.

The functional similarity in facial motion selectivity between monkey and human STS regions demonstrated in the current study is in line with previous comparisons of cortical function between the two species. For example, some studies have reported striate and

extrastriate visual area homologues using electrophysiological recordings in monkeys (Sereno et al., 1994) and fMRI in humans (Sereno et al., 1995), while others have reported cross-species homologies in the ventral visual pathway for face and place selectivity (Lafer-Sousa and Conway, 2013; Lafer-Sousa et al., 2016). Thus our results naturally extend this homology to the functional specialization of the superior temporal sulcus in both species for processing facial motion.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Intramural Research Program of the National Institute of Mental Health (ZIAMH002918; NCT00001360), and by National Natural Science Foundation of China under Grant No. 81871511. We thank Frank Ye, Charles Zhu, and David Yu for technical assistance during monkey scans, and Kenny Kan for technical assistance during human scans.

References

Allison T, Puce A, McCarthy G, 2000 Social perception from visual cues: role of the STS region. Trends Cognit. Sci 4, 267–278. [PubMed: 10859571]

Aggarwal JK, Cai Q, Liao W, Sabata B, 1998 Nonrigid motion analysis: articulated and elastic motion. Comput. Vis. Image Understand. 70 (2), 142–156.

Ambadar Z, Schooler JW, Cohn JF, 2005 Deciphering the enigmatic face: the importance of facial dynamics in interpreting subtle facial expressions. Psychol. Sci 16, 403–410. [PubMed: 15869701]

Andrews TJ, Ewbank MP, 2004 Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. Neuroimage 23, 905–913. [PubMed: 15528090]

Axelrod V, Yovel G, 2013 The challenge of localizing the anterior temporal face area: a possible solution. Neuroimage 81, 371–380. [PubMed: 23684864]

Baker S, Matthews I, 2004 Lucas-Kanade 20 years on: a unifying framework. IJCV 53 (3), 221–255.

Beauchamp MS, 2015 The social mysteries of the superior temporal sulcus. Trends Cognit. Sci 19 (9), 489–490. [PubMed: 26208834]

Beauchamp MS, Lee KE, Argall BD, Martin A, 2004 Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41 (5), 809–823. [PubMed: 15003179]

Beauchamp MS, Lee KE, Haxby JV, Martin A, 2002 Parallel visual motion processing streams for manipulable objects and human movements. Neuron 34 (1), 149–159. [PubMed: 11931749]

Beauchamp MS, Lee KE, Haxby JV, Martin A, 2003 FMRI responses to video and point-light displays of moving humans and manipulable objects. J. Cognit. Neurosci 15 (7), 991–1001. [PubMed: 14614810]

Beauchemin SS, Barron JL, 1995 The computation of optical flow. ACM Comput. Surv 27, 433–466.

Bell AH, Hadj-Bouziane F, Frihauf JB, Tootell RB, Ungerleider LG, 2009 Object representations in the temporal cortex of monkeys and humans as revealed by functional magnetic resonance imaging. J. Neurophysiol 101, 688–700. [PubMed: 19052111]

Bernstein M, Yovel G, 2015 Two neural pathways of face processing: a critical evaluation of current models. Neurosci. Biobehav. Rev 55, 536–546. [PubMed: 26067903]

Bonda E, Petrides M, Ostry D, Evans A, 1996 Specific involvement of human parietal systems and the amygdala in the perception of biological motion. J. Neurosci 16, 3737–3744. [PubMed: 8642416]

Bruce V, Young A, 1990 Understanding face recognition. Br. J. Psychol 81, 361–380. Comment in. [PubMed: 2224396]

Bruhn A, Weickert J, Schnorr C, 2005 Lucas/Kanade meets Horn/Schunck: combining local and global optic flow methods. Int. J. Comput. Vis 61, 211–231.

Cox RW, 1996 AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res 29, 162–173. [PubMed: 8812068]

Deen B, Koldewyn K, Kanwisher N, Saxe R, 2015 Functional organization of social perception and cognition in the superior temporal sulcus. Cerebr. Cortex 25 (11), 4596–4609.

Desimone R, Ungerleider LG, 1986 Multiple visual areas in the caudal superior temporal sulcus of the macaque. J. Comp. Neurol 248, 164–189. [PubMed: 3722457]

Dubner R, Zeki SM, 1971 Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. Brain Res 35, 528–532. [PubMed: 5002708]

Duchaine B, Yovel G, 2015 A revised neural framework for face processing. Annu. Rev. Vis. Sci 1, 293–416.

Fisher C, Freiwald WA, 2015a Contrasting specializations for facial motion within the macaque face-processing system. Curr. Biol 25, 261–266. [PubMed: 25578903]

Fisher C, Freiwald WA, 2015b Whole-agent selectivity within the macaque face-processing system. Proc. Natl. Acad. Sci. U. S. A 112, 14717–14722. [PubMed: 26464511]

Fleet DJ, Weiss Y, 2006 Optical flow estimation In: Paragios et al. Handbook of Mathematical Models in Computer Vision. Springer, pp. 239–257.

Fox CJ, Iaria G, Barton JJ, 2009 Defining the face-processing network: optimization of the functional localizer in fMRI. Hum. Brain Mapp 30, 1637–1651. [PubMed: 18661501]

Furl N, Hadj-Bouziane F, Liu N, Averbeck BB, Ungerleider LG, 2012 Dynamic and static facial expressions decoded from motion-sensitive areas in the macaque monkey. J. Neurosci 32, 15952–15962. [PubMed: 23136433]

Hadj-Bouziane F, Bell AH, Knusten TA, Ungerleider LG, Tootell RB, 2008 Perception of emotional expressions is independent of face selectivity in monkey inferior temporal cortex. Proc. Natl. Acad. Sci. U.S.A 105, 5591–5596. [PubMed: 18375769]

Hadj-Bouziane F, Liu N, Bell AH, Gothard KM, Luh WM, Tootell RB, Murray EA, Ungerleider LG, 2012 Amygdala lesions disrupt modulation of functional MRI activity evoked by facial expression in the monkey inferior temporal cortex. Proc. Natl. Acad. Sci. U. S. A 109, E3640–E3648. [PubMed: 23184972]

Haxby JV, Hoffman EA, Gobbini MI, 2000 The distributed human neural system for face perception. Trends Cognit. Sci 4, 223–233. [PubMed: 10827445]

Harris RJ, Young AW, Andrews TJ, 2012 Morphing between expressions dissociates continuous from categorical representations of facial expression in the human brain. Proc. Natl. Acad. Sci. U.S.A 109, 21164–21169. [PubMed: 23213218]

Herrington JD, Nymberg C, Schultz RT, 2011 Biological motion task performance predicts superior temporal sulcus activity. Brain Cognit 77, 372–381. [PubMed: 22024246]

Jastorff J, Popivanov ID, Vogels R, Vanduffel W, Orban GA, 2012 Integration of shape and motion cues in biological motion processing in the monkey STS. Neuroimage 60 (2), 911–921. [PubMed: 22245356]

Kanwisher N, McDermott J, Chun MM, 1997 The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci 17, 4302–4311. [PubMed: 9151747]

Kanwisher N, Yovel G, 2006 The fusiform face area: a cortical region specialized for the perception of faces. Phil. Trans. Roy. Soc. Lond. B 361, 2109–2128. [PubMed: 17118927]

Knappmeyer B, Thornton IM, Bülthoff HH, 2003 The use of facial motion and facial form during the processing of identity. Vis. Res 43, 1921–1936. [PubMed: 12831755]

Knight B, Johnston A, 1997 The role of movement in face recognition. Vis. Cognit 4, 265–273.

Kriegeskorte N, Formisano E, Sorger B, Goebel R, 2007 Individual faces elicit distinct response patterns in human anterior temporal cortex. Proc. Natl. Acad. Sci. U.S.A 104, 20600–20605. [PubMed: 18077383]

Lafer-Sousa R, Conway BR, 2013 Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. Nat. Neurosci 16 (12), 1870–1878. [PubMed: 24141314]

Lafer-Sousa R, Conway BR, Kanwisher NG, 2016 Color-biased regions of the ventral visual pathway lie between face- and place-selective regions in humans, as in macaques. J. Neurosci 36 (5), 1682–1697. [PubMed: 26843649]

Lander K, Christie F, Bruce V, 1999 The role of movement in the recognition of famous faces. Mem. Cognit 27, 974–985.

Liu N, Kriegeskorte N, Mur M, Hadj-Bouziane F, Luh WM, Tootell RB, Ungerleider LG, 2013 Intrinsic structure of visual exemplar and category representations in macaque brain. J. Neurosci 33, 11346–11360. [PubMed: 23843508]

Lucas BD, Kanade T, 1981. An iterative image registration technique with an application to stereo vision. In: Proc 7th Intl Joint Conf on Artificial Intelligence (IJCAI), 2, pp. 674–679, 1981.

McCarthy G, Puce A, Gore JC, Allison T, 1997 Face-specific processing in the human fusiform gyrus. J. Cognit. Neurosci 9, 605–610. [PubMed: 23965119]

McMahon DB, Russ BE, Elnaiem HD, Kurnikova AI, Leopold DA, 2015 Single-unit activity during natural vision: diversity, consistency, and spatial sensitivity among AF face patch neurons. J. Neurosci 35, 5537–5548. [PubMed: 25855170]

O'Toole AJ, Roark DA, Abdi H, 2002 Recognizing moving faces: a psychological and neural synthesis. Trends Cognit. Sci 6, 261–266. [PubMed: 12039608]

Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N, 2011 Differential selectivity for dynamic versus static information in face-selective cortical regions. Neuroimage 56, 2356–2363. [PubMed: 21473921]

Polosecki P, Moeller S, Schweers N, Romanski LM, Tsao DY, Freiwald WA, 2013 Faces in motion: selectivity of macaque and human face processing areas for dynamic stimuli. J. Neurosci 33, 11768–11773. [PubMed: 23864665]

Puce A, Allison T, Bentin S, Gore JC, McCarthy G, 1998 Temporal cortex activation in humans viewing eye and mouth movements. J. Neurosci 18, 2188–2199. [PubMed: 9482803]

Roark DA, O'Toole AJ, Abdi H, Barrett SE, 2006 Learning the moves: the effect of familiarity and facial motion on person recognition across large changes in viewing format. Perception 35, 761–773. [PubMed: 16836043]

Schall S, Kiebel SJ, Maess B, von Kriegstein K, 2015 Voice identity recognition: functional division of the right STS and its behavioral relevance. J. Cognit. Neurosci 27 (2), 280–291. [PubMed: 25170793]

Schultz J, Brockhaus M, Bülthoff HH, Pilz KS, 2013 What the human brain likes about facial motion. Cerebr. Cortex 23, 1167–1178.

Schultz J, Pilz KS, 2009 Natural facial motion enhances cortical responses to faces. Exp. Brain Res 194, 465–475. [PubMed: 19205678]

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB, 1995 Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268, 889–893. [PubMed: 7754376]

Sereno MI, McDonald CT, Allman JM, 1994 Analysis of retinotopic maps in extrastriate cortex. Cerebr. Cortex 4, 601–620.

Sliwa J, Freiwald WA, 2017 A dedicated network for social interaction processing in the primate brain. Science 356 (6339), 745–749. [PubMed: 28522533]

Trautmann SA, Fehr T, Herrmann M, 2009 Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. Brain Res 1284, 100–115. [PubMed: 19501062]

Tsao DY, Moeller S, Freiwald WA, 2008 Comparing face patch systems in macaques and humans. Proc. Natl. Acad. Sci. U. S. A 105, 19514–19519. [PubMed: 19033466]

Wehrle T, Kaiser S, Schmidt S, Scherer KR, 2000 Studying the dynamics of emotional expression using synthesized facial muscle movements. J. Pers. Soc. Psychol 78, 105–119. [PubMed: 10653509]

Willems RM, Peelen MV, Hagoort P, 2010 Cerebral lateralization of face-selective and body-selective visual areas depends on handedness. Cerebr. Cortex 20, 1719–1725.

Yovel G, Tambini A, Brandman T, 2008 The asymmetry of the fusiform face area is a stable individual characteristic that underlies the left-visual-field superiority for faces. Neuropsychologia 46, 3061–3068. [PubMed: 18639566]

Zhu Q, Nelissen K, Van den Stock J, De Winter FL, Pauwels K, de Gelder B, Vanduffel W, Vandenbulcke M, 2013 Dissimilar processing of emotional facial expressions in human and monkey temporal cortex. Neuroimage 66, 402–411. [PubMed: 23142071]

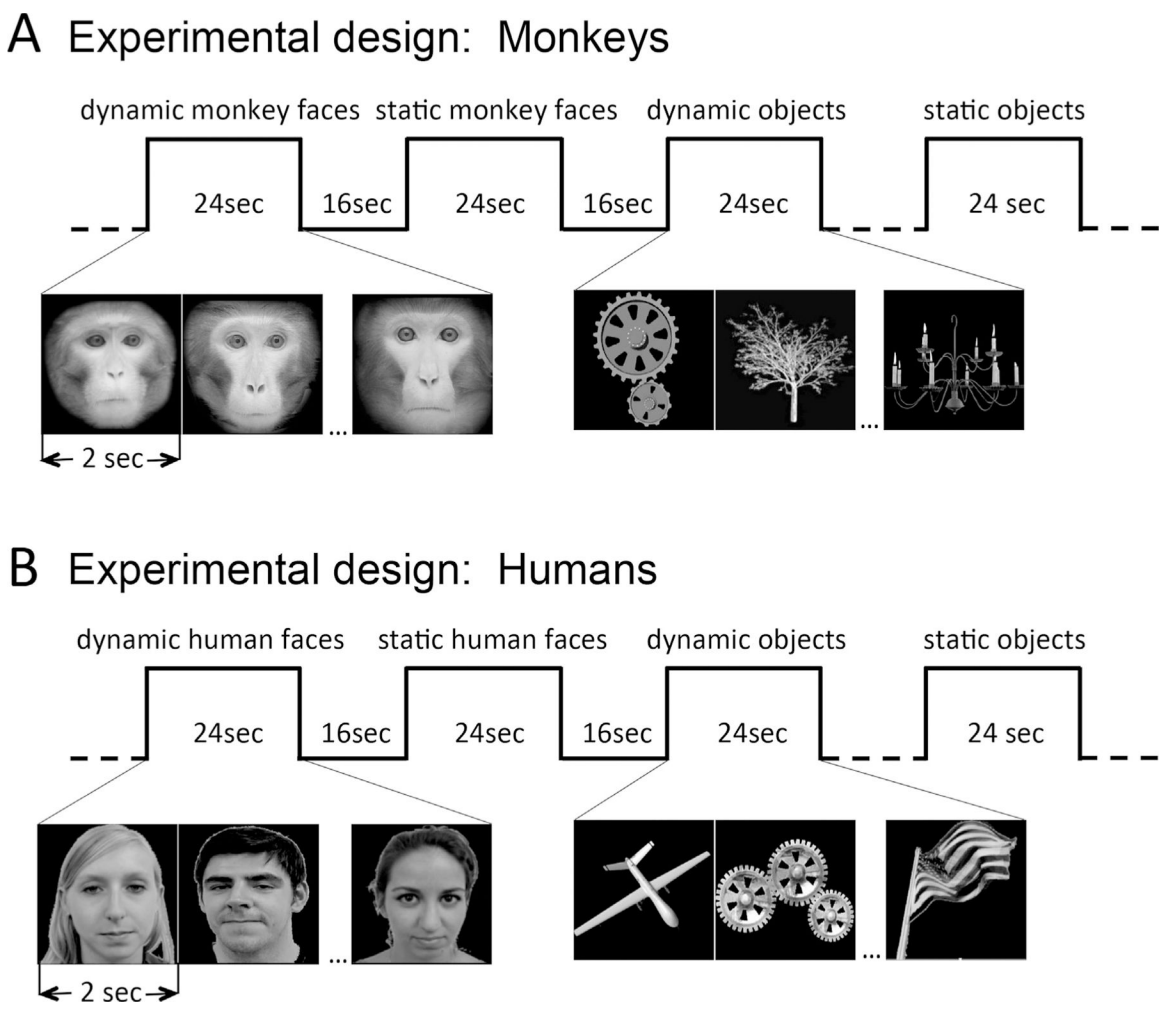

Fig. 1.

A) Monkeys viewed blocks of dynamic monkey faces, dynamic objects, static monkey faces and static objects in the main task. B) Humans viewed blocks of dynamic human faces, dynamic objects, static human faces and static objects in the main task. Each dynamic condition contained 12 different video clips, and each static condition contained 12 different images that were generated from the corresponding dynamic video clip. Each block lasted 24 s, with 16-sec baseline fixation periods between blocks. Within a block, each video clip (for dynamic conditions) or each image was presented for 2 s, with no blank periods between them. Motion energy in the dynamic face video clips was equivalent (or as close as possible) to that in the corresponding dynamic object videos for each block of the stimuli using an optic-flow algorithm.

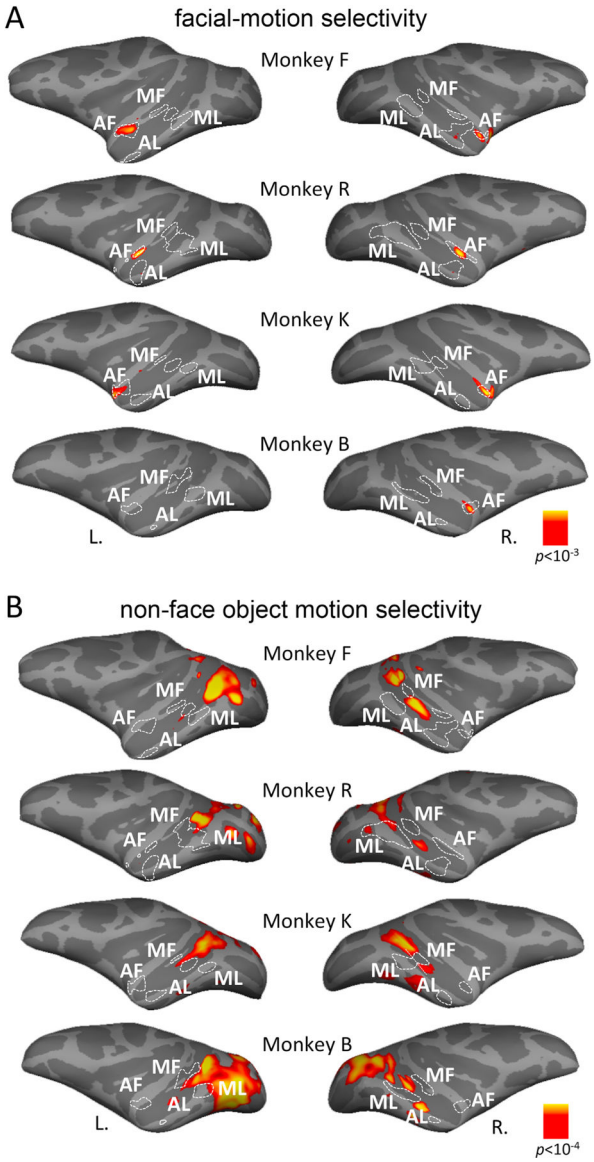

Fig. 2.

A) Localization of facial motion selective regions (patches) by contrasting the fMRI response to motion caused by faces relative to motion caused by objects [(dynamic faces > static faces) - (dynamic objects > static objects)] in monkeys F, R, K and B. The location of these facial motion patches overlapped the AF face patches, for monkeys F, R and K bilaterally, and for monkey B in the right hemisphere. B) Localization of object motion selective regions by contrasting the fMRI response to motion caused by non-face objects relative to motion caused by faces [(dynamic objects > static objects) - (dynamic faces > static faces)] in each of the four monkeys. The object motion selective regions are mostly in the posterior and middle portions of STS in all 4 monkeys. Dashed white lines outline areas of static face-selective regions shown in Supplementary Figure S4A.

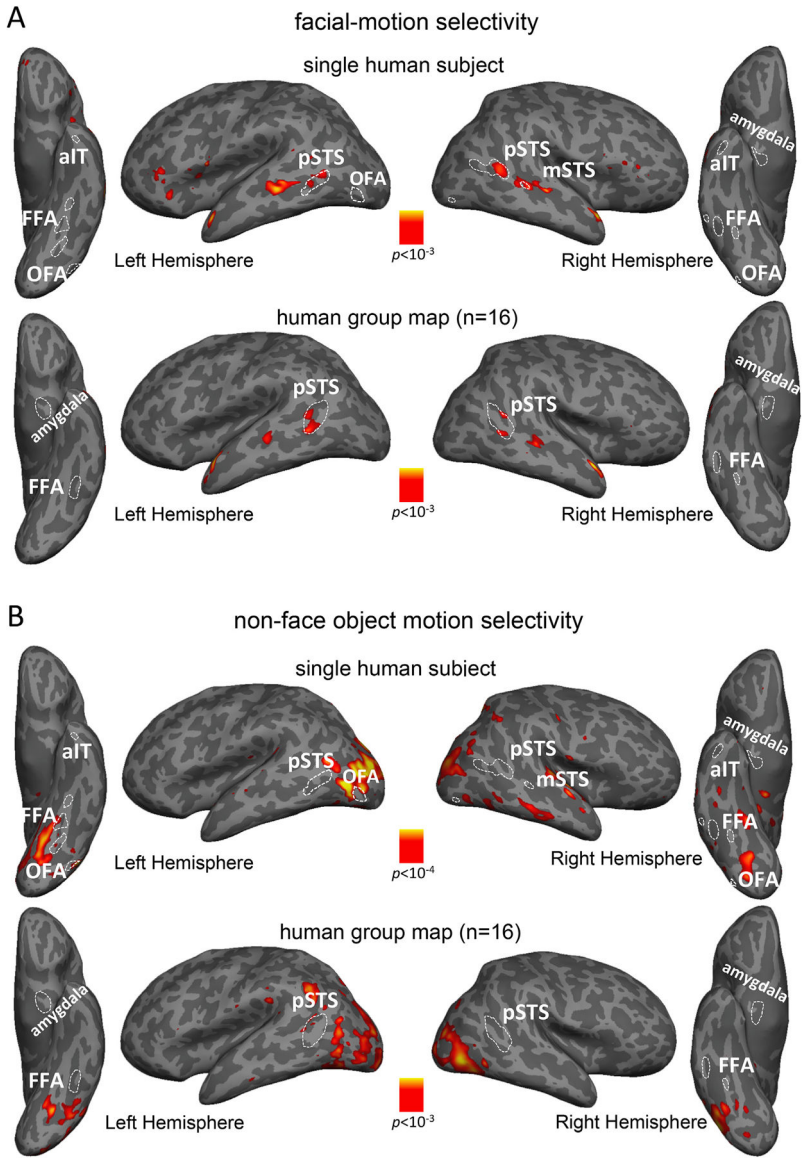

Fig. 3.

A) Localization of facial motion selective regions for a representative human subject by contrasting the fMRI response to motion caused by faces relative to motion caused by objects. This contrast identified three separate foci along the STS: anterior STS, middle STS and posterior STS. The posterior STS overlapped with the pSTS face-selective region. B) Localization of object motion selective regions for the representative human subject by contrasting the fMRI response to motion caused by objects relative to motion caused by faces. The identified object motion selective regions are mostly outside of human STS. Dashed white lines outline areas of static face-selective regions shown in Supplementary Figure S5A.

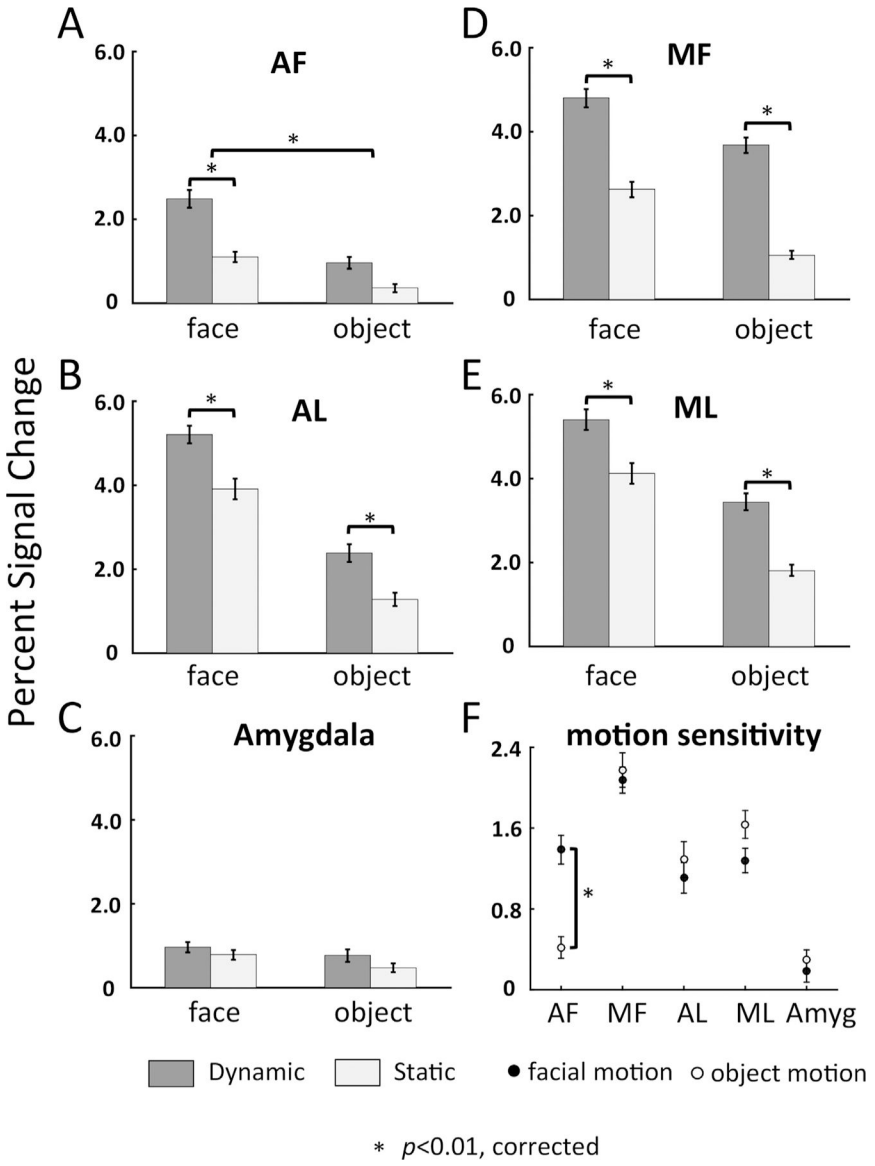

Fig. 4.

A-E) Percent signal change of the fMRI response amplitude for each category of visual stimuli (dynamic faces, static faces, dynamic objects, static objects) in the monkey STS regions-of-interest. AF: Anterior Fundus, MF: Middle Fundus, AL: Anterior Lateral, ML: Middle Lateral. F) The facial motion and object motion sensitivity for each monkey ROI. * indicate $p < 0.01$ (Bonferroni corrected).

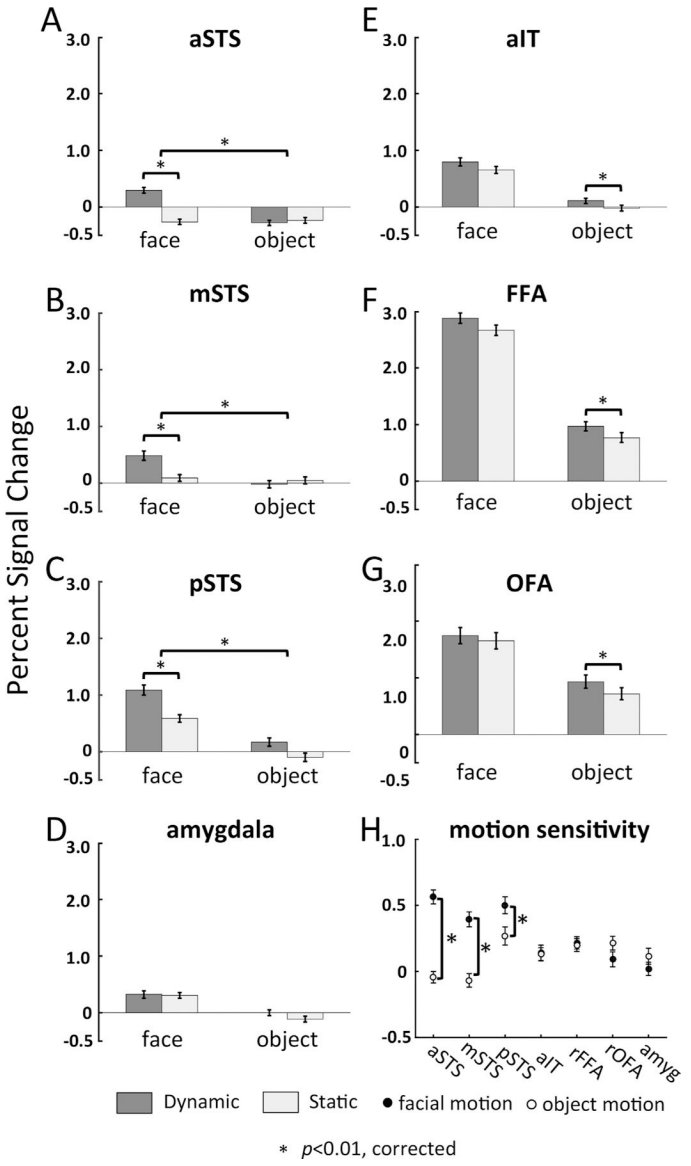

Fig. 5.

A-G) Percent signal change of the fMRI response amplitude for each category of visual stimuli (dynamic faces, static faces, dynamic objects, static objects) in the human regions-of-interest. aSTS: anterior Superior Temporal Sulcus, mSTS: middle Superior Temporal Sulcus, pSTS: posterior Superior Temporal Sulcus, aIT: anterior Inferior Temporal cortex, FFA: Fusiform Face Area, OFA: Occipital Face Area. H) The facial motion and object motion sensitivity for each human ROI. * indicate $p < 0.01$ (Bonferroni corrected).

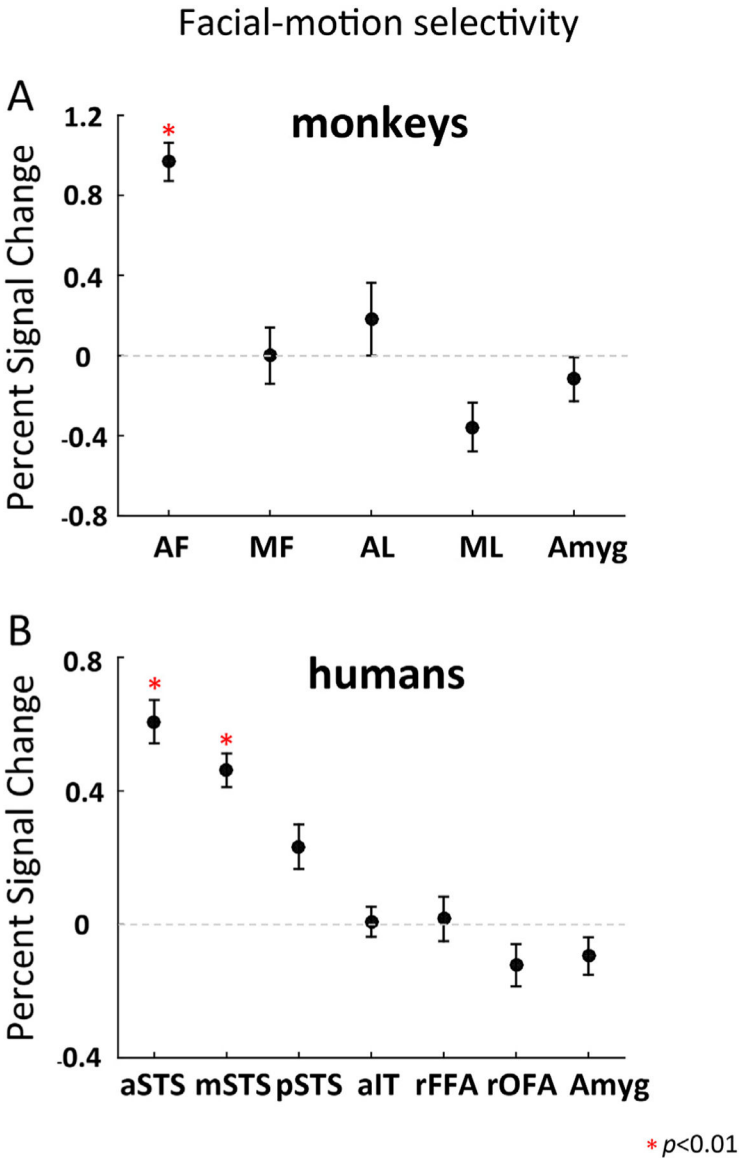

Fig. 6.

The PSC values of facial motion selectivity in the regions of: A) monkeys and B) humans. In monkeys, only AF showed significantly greater facial motion selectivity than zero. In humans, only aSTS and mSTS showed significantly greater facial motion selectivity than zero. * indicate $p < 0.01$ (Bonferroni corrected).

Table 1

Activation peaks for monkey facial-motion selective regions registered to monkey D99 template.

Region		Saleem and Logothetis coordinates (x, y, z)/Cluster size (mm^3)			
		Monkey F	**Monkey R**	**Monkey K**	**Monkey B**
anterior fundus STS	L.	(−18, 20, 2)/30.2	(−21, 20, −3)/18.4	(−18, 22, −4)/22.2	–
	R.	(18, 20, −2)/35.6	(20, 19, −2)/18.7	(19, 19, −2)/26.6	(20, 24, −2)/12.1

Table 2

Averaged activation peaks for human facial-motion selective regions registered to Talairach template.

Region		Talairach Peak Coordinates (Mean±SD)			Number of Subjects	Cluster size (mm^3) (Mean±SD)
		x	y	z		
aSTS	L.	−54±5.1	6±4.7	−15±4.5	13	64±35
	R.	57±4.7	8±4.6	−16±4.6	16	113±46
mSTS	L.	−48±4.8	−18±5:7	−2±3.3	11	77±42
	R.	48±4.4	−13±4.6	−8±4.5	16	95±45
pSTS	L.	−52±5.5	−41±5.2	4±5.1	16	185±52
	R.	52±4.6	−38±4.9	2±4.8	16	204±56